

Draft benchmark description: FluidFlower international benchmark study

Jan M. Nordbotten¹, Martin Fernø², Bernd Flemisch³, Ruben Juanes⁴, Magne Jørgensen⁵

--- DRAFT VERSION 2.3 ---

Update notes from Version 2.2:

Chapter 2.2	Updated Figures 5- 8 geometry, changed dimensions of Box B
Chapter 2.2.2	Description of pre-injection flushing + image of compaction
Chapter 2.3	Description on absolute permeability, endpoint relative perm and entry pressures. Results from measurements
Chapter 2.7	Well tests – description changed
Appendix A	Grain size distributions, images, graphs and histograms
Appendix B	Well test images, pressure profiles

Abstract

Numerical benchmark studies of multi-phase flow in porous media generally lack a clearly defined ground truth. In this benchmark study, we consider a problem inspired by CO₂ storage, where the dominant processes are associated with multiphase flows, capillarity, dissolution, and convective mixing. Supporting the numerical benchmark study, we conduct a series of physical realizations in the FluidFlower experimental setup.

1. Overview and vision

The group of Professor Fernø at the University of Bergen has constructed a new experimental rig, named “FluidFlower”, as shown in Figure 1. The rig allows for relatively large-scale multiphase 2D flow experiments (almost 3 by 2 meters) on model geological geometries, with an unprecedented data acquisition in terms of a pressure sensor network (seen in the background) together with optical registration of both saturation and acidity (pH). Moreover, the design allows for repeated experiments within the exact same experimental setup, allowing physical uncertainty to be addressed (see also: <https://fluidflower.w.uib.no/>).

¹ Department of Mathematics, University of Bergen

² Department of Physics and Technology, University of Bergen

³ Institut für Wasser- und Umweltsystemmodellierung, Universität Stuttgart

⁴ Department of Civil and Environmental Engineering, Massachusetts Institute of Technology,

⁵ Simula Research Laboratory

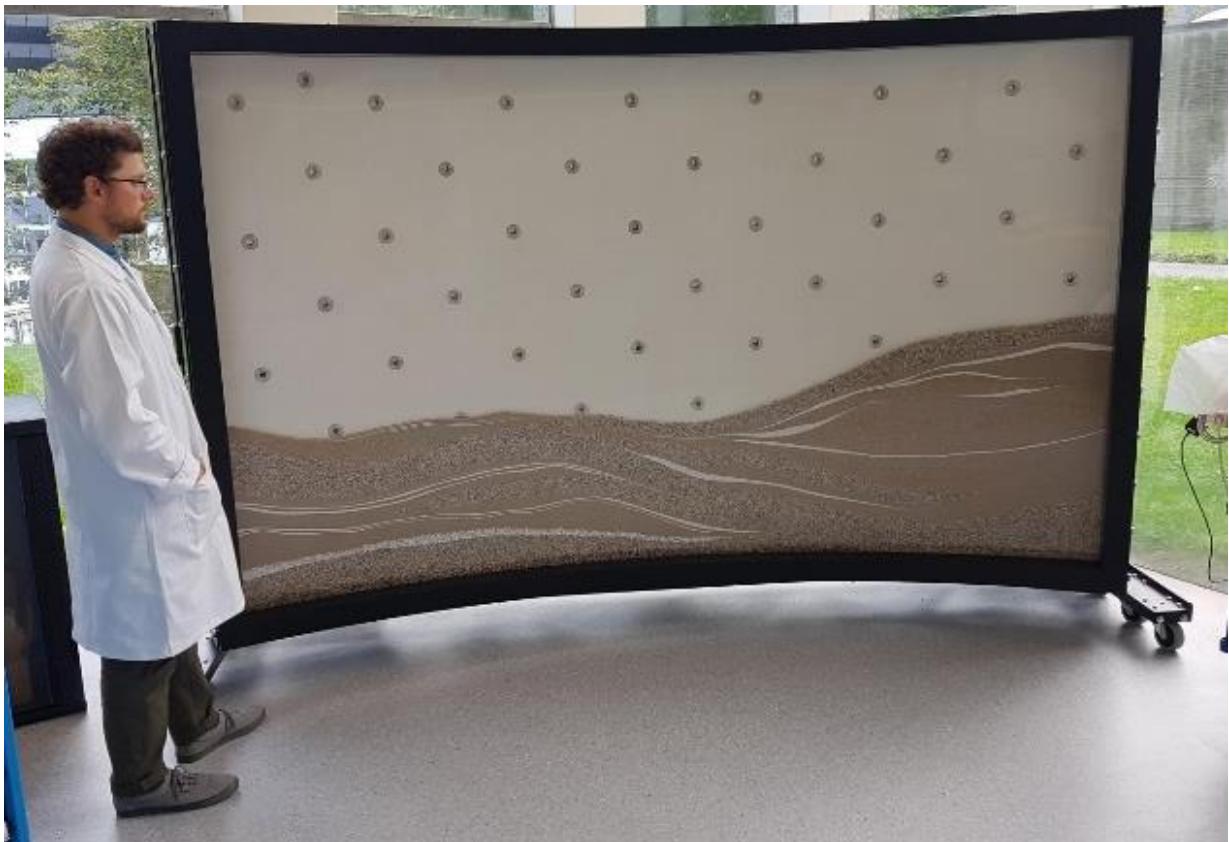


Figure 1. Full-scale FluidFlower rig, half-filled with geological layers (lower part). In the upper part, which is not yet filled in this photo, the pressure sensor network is visible.

1.1 Benchmark study

To leverage these new experimental capabilities, we are conducting this international benchmarking study for modeling, simulation and forecasting of multiphase, multicomponent flows in porous media. Several multiphase flow phenomena are present in the experimental setup, including both heterogeneous and hysteretic two-phase flow properties (leading to both structural and residual gas trapping) and two-component mixing with concomitant development of gravity fingers, as illustrated in a sample geometry in Figure 2.

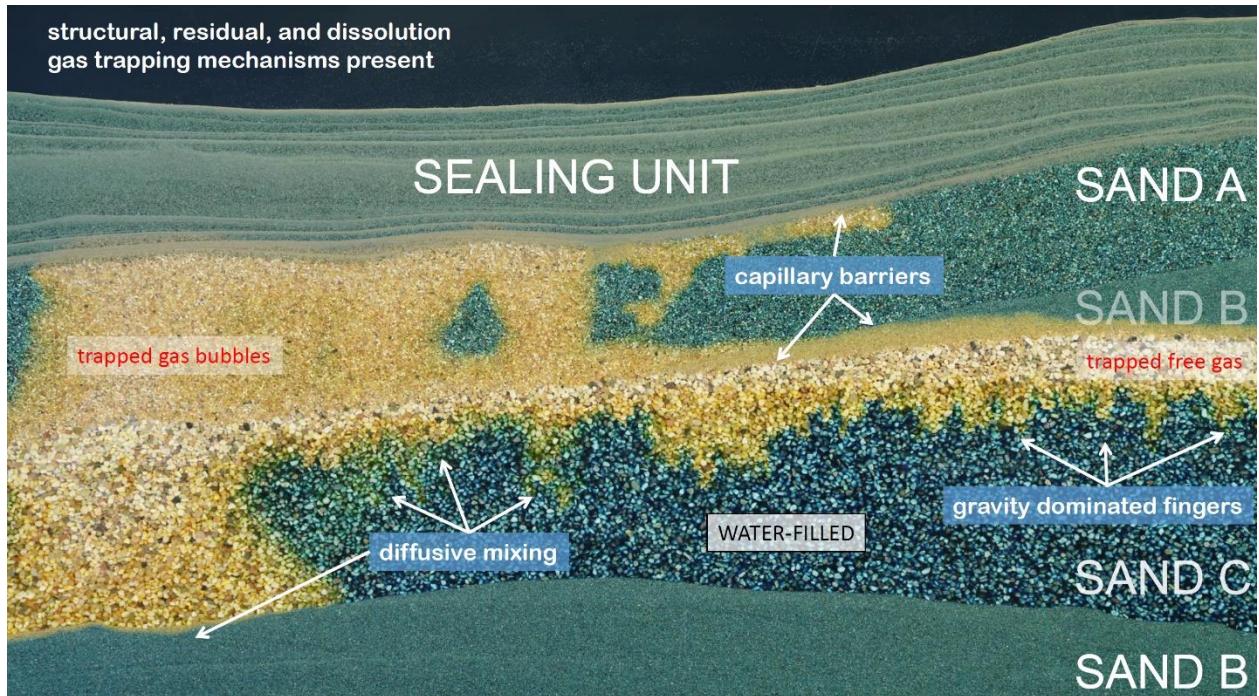


Figure 2: Sample experiment in a downscaled FluidFlower rig. Visible here are multiple fluid-filled gas layers, free-phase CO₂, and variable concentrations of dissolved CO₂ in the water phase. Important processes for CO₂ storage, such as capillary entry effects, hysteresis, gravity fingering, are all apparent.

The main goal of the benchmark is to provide a full-physics validation of the state-of-the-art simulation capabilities within the international porous media community. To this aim we envision the following benchmarking process:

International numerical benchmark study:

- A full geometric, operational, and petrophysical description is hereby provided to all participants, including commonly measured multiphase flow parameters for porous media simulation based on *ex situ* measurements (grain size distributions, porosities, permeabilities, end-point relative permeabilities, entry pressures). All parameters are given with our best estimates of tolerances.
- Additionally, experimental single-phase flow data (well tests and tracer tests) are provided to the participants for model calibration.
- Participants must themselves provide equation-of-state data, based on the liquid and gas used.
- All participants are asked to provide simulation output in terms of target quantities, together with both quantitative and qualitative confidence intervals, according to a common protocol as specified in Section 3.

Physical experiments:

- On the exact geometry provided to the participants, UoB will conduct at least 5 identical multi-phase flow experiments.

In-house parameter estimation at NORCE and MIT:

- Complementing the benchmarking study, NORCE will conduct an in-house numerical parameter estimation study based on the full set of multi-phase flow experiments.
- Also complementing the benchmarking study, MIT and UoB will collaborate on assessing the value of using an analog system (here in the form of a smaller version of the FluidFlower) for calibrating numerical simulations.

Together, the physical experiments and numerical simulations will allow us to address the confidence with which the main physical processes associated with subsurface CO₂ storage can be modeled:

- A) The physical experiments will provide both a “ground truth”, and will also allow us to quantify to some extent the inherent uncertainties associated with the porous media flow itself by means of repeated experiments on the same geometry.
- B) The in-house parameter estimation study will allow us to address any discrepancies between *ex situ* measured parameters and *in situ* properties, as well as to provide a sense of the extent the standard model equations for porous media can capture the observed flows.
- C) The numerical simulations reported by each research group will allow us to quantify the precision and intra-group uncertainty as assessed by each of the research groups individually.
- D) The total panel of research groups will allow us to quantify the inter-group variability in the results (and thus to some extent the reproducibility of numerical modeling), and the relationship between the ensemble precision and uncertainty.

The combination of A), B), C) and D) within the same benchmarking study is to our knowledge unprecedented within the porous media community. This will allow us, within the context of this setup, to address for the first time the correlation between the numerical uncertainty quantification and a measure of real system uncertainty.

[1.2 Participants](#)

Following a call for participation, the following nine research groups have joined the benchmark study:

Institution	Persons
UT Austin	Nick Espinosa, Mohamad Jammoul, Mary Wheeler
CSIRO	Jonathan Ennis-King, James Gunning, Samuel Jackson, Andy Wilkins
TU Delft	Hadi Hajibeygi, Denis Voskov, Michiel Wapperom
Herriot-Watt	Florian Doster, Sebastian Geiger
Imperial College	Matthew Jackson
LANL	Bill Carey, Satish Karra, Hari Viswanathan
Melbourne	Stephan Matthai, Qi Shao, AbdAllah Youssef
Stanford	Jacques Franc, Hamdi Tchelepi
Stuttgart	Holger Class, Dennis Gläser, Fabian Jobst

In addition, Equinor ASA (represented by Ali Mojaddam Zadeh and Bamshad Nazarian), have expressed interest in participating in the benchmark through their collaboration with LANL.

1.3 Time frame

- 30. April, 2021: Closed call for participation opens.
- 15. June, 2021: Call closes.
- 15. July, 2021: Preliminary benchmark description supplied to participants.
- 16. July – 19. August, 2021: Preparation phase, discussion possible among all participants and the experimental team.
- 20. August, 2021: Deadline for feedback on preliminary benchmark description.
- 16. September, 2021: Kick-off Zoom meeting, second iteration of benchmark description distributed.
- 17. September – 8. October, 2021: Open discussion for finalizing the description.
- 30. September, 2021: Submission of signed participation agreements.
- 8. October, 2021: Final benchmark description circulated to participants.
- 9. October – 14. December, 2021: Blind phase, no direct communication between different participants or with the experimental group.
- 12. December, 2021: **Deadline for submitting blind benchmark data.**
- 15. December, 2021: Virtual workshop and comparison of “fully blind” simulation forecasts.
- 15. December, 2021 – 13. March, 2022: Synchronization phase, communication between all participants enabled, but not with the experimental group.
- 10. March, 2022: **Deadline for submitting final benchmark data.**
- 2 days in the week 14.-18. March, 2022: Real workshop in Norway (location to be determined) with presentation of final simulation forecasts, experimental results, model calibration study, and synthesis of results.
- Spring 2022: Writing and submitting papers based on the findings of the study.

2. Benchmark description

The geometric description is motivated by typical North Sea reservoirs, and has been developed in consultation with faculty and researchers at the Department of Geology, UoB. A special thanks goes to *Robert Gawthorpe, Atle Rotlevatn and Casey Nixon for their helpful comments*. The physical and petrophysical properties are based on unconsolidated sands, as measured by the group of Professor Martin Fernø at the University of Bergen.

2.1 External Geometry

The porous media in the FluidFlower is contained between an optically transparent front panel and a sealed back panel with perforations for fluid injection/production and pressure measurements. The length of the porous media is 2840 mm and height is 1300 ± 30 mm (note large \pm due to non-horizontal top porous media surface, see Figure 6 for exact height). Note that the visible length of the porous media is 2800 mm, but the active porous media extends 3 cm behind black metal frame on each side). The porous media is curved (both the front and back panels are curved to be able to sustain internal forces of sand + liquids); the curvature is 1/8 of a circle with radius 3.6 m. The nominal width of the porous media between the front and back panel is 19 mm, but varies both from left-to-right and top-to-bottom due to internal forces when filled with sand and water. The width at each side (far left and far right) is fixed at 19 mm, and increases towards the center to 28 mm, see Figure 3. Note that the widths have been measured after initial sand filling of the rig by removing the front panel. The height of the coarse-grained sand used (not used in benchmark) was measured point-wise along the left-to-right and top-to-bottom axes.



Figure 3: Porous media width variation (in mm) within the curved flow rig, using an image corrected curvature. The width at each side (far left and far right) is fixed at 19 mm. The laser grid (red squares) is 100 x 100 mm. Injection ports (in red circles) and pressure measurement ports (purple) are described in more detail below.

The FluidFlower has 56 perforations (see Figure 4 for description) for fluid injection and pressure measurements. During the benchmark study there will be two injection ports and two pressure ports. The remaining perforations will be closed (some are permanently plugged, some are able to serve as additional pressure/injection/outlet ports). The four open ports (pressure measurements and fluid injection) have an inner diameter of 1.8 mm. The two injection ports extend 10 mm out from the back panel (a nominal outer diameter of extension tube is 1/8 in (3.2 mm)) into the sand to reduce gas flow along the back. All other ports (except bottom row) are flush with the back panel, and are sealed with a custom-made plastic (POM) port. A thin layer of silicone is used to maintain the seal to avoid leakage. Upon injection of CO₂ gas, excess fluids will be produced from the top row from several ports with 10+ mm diameter.

The bottom row (blue ports, see Figure 4) are technical ports for re-setting fluids between CO₂ injection tests, and will remain closed during CO₂ injections. Each of the seven blue ports have stainless steel tubes (same inner and outer diameter as above) that point downwards into the lowest (high-permeable) sand layer and stop approximately 5 mm above the bottom of the porous media. The steel tubes extend 8 mm out from the back panel, and the space between the tube and back panel is filled with silicone.

The FluidFlower boundaries are closed at the bottom and both left and right sides (no-flow boundaries). The top is open and in contact with (fluctuating) atmospheric pressure, with a free water table fixed at a constant elevation (constant hydraulic head) of 1500 mm above the bottom. Ports in the top row are alternating production/injection ports to ensure a fixed water table.

2.2 Geology

Figure 4 provides a sketch of the intended geometry, whereas Figure 5 is the actual geometry. To avoid discrepancies between the intended and actual geometries, the actual geometry is provided as a high-resolution photograph, shown here in reduced resolution as Figure 5.

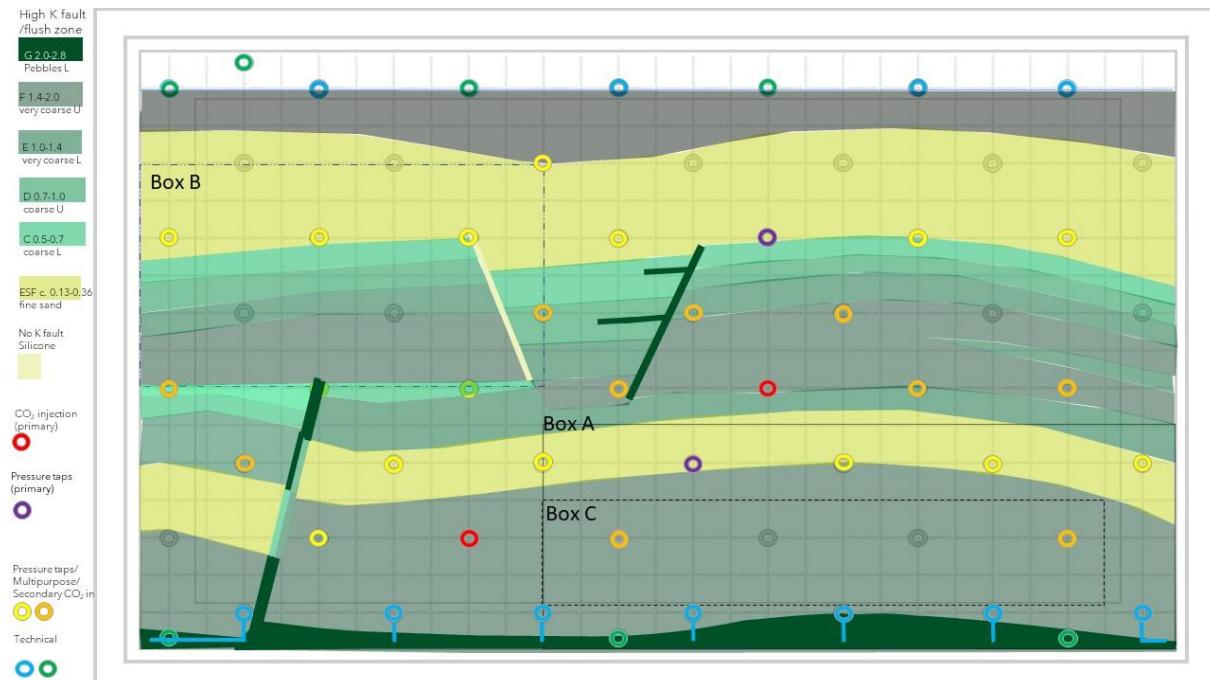


Figure 4: A sketch of the intended benchmark geometry. The geometry includes an anticline (right side) where CO₂ accumulation is anticipated. There are three fault-like structures in the geometry, with different permeability (two high and one low zero). The lower fault (left side) is "heterogeneous" and

consists of several sand types. This approach is used to reduce smaller grained sands leaking into large-grained sands. The upper, left fault is sealed (using silicone) and may be regarded as a sealing fault. The upper right fault is high-permeable and built using a single sand type. There are two CO₂ injection points (red circles) and two pressure taps (purple circles). Box A, B and C are indicated in the figure, and described in more detail below. Each gray square in the grid in the background is 10 x 10 cm. A *high-resolution version of sketch is available for the benchmark study, and is available on GitHub (<https://github.com/fluidflower>).*



Figure 5: A photograph image of the actual benchmark geometry in the FluidFlower rig. The image is corrected for curvature using Matlab. Note that the silicone in the sealed fault is red due to acetic acid compounds reacting with the blue-green pH indicator solution. *The high-resolution version of this data is considered the most accurate geological description available for the benchmark study, and is made available at (<https://github.com/fluidflower>).*

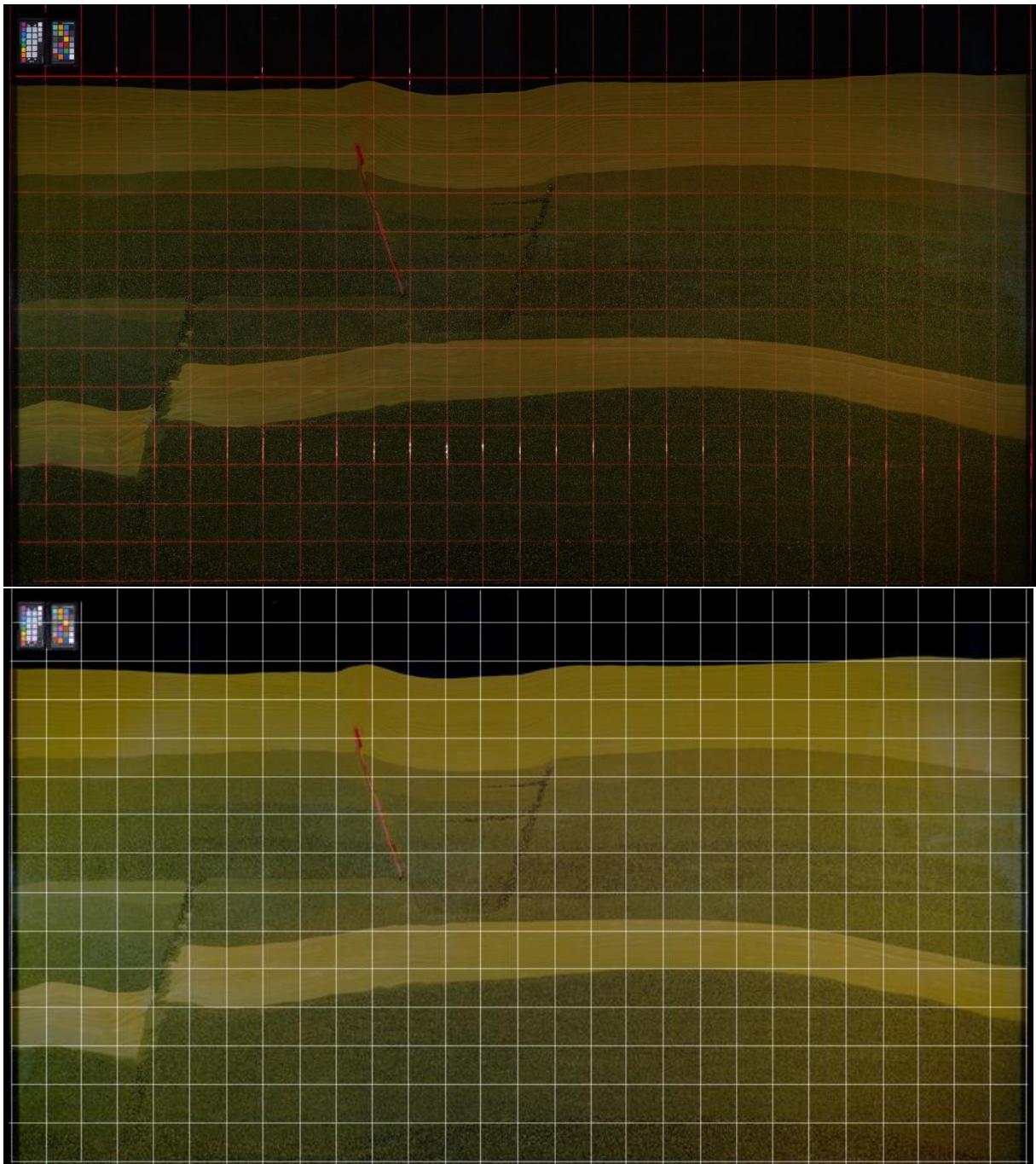


Figure 6: Top: A photograph image of the benchmark geometry with laser grid (each square is 100 x 100 mm to within +/- 1 mm). The image is corrected for curvature using an inhouse Matlab script. Bottom: enhanced grid to ease interpretation. It is important to note that the laser grid is considered the most accurate measure of the geometry, and that the image may still be slightly distorted.

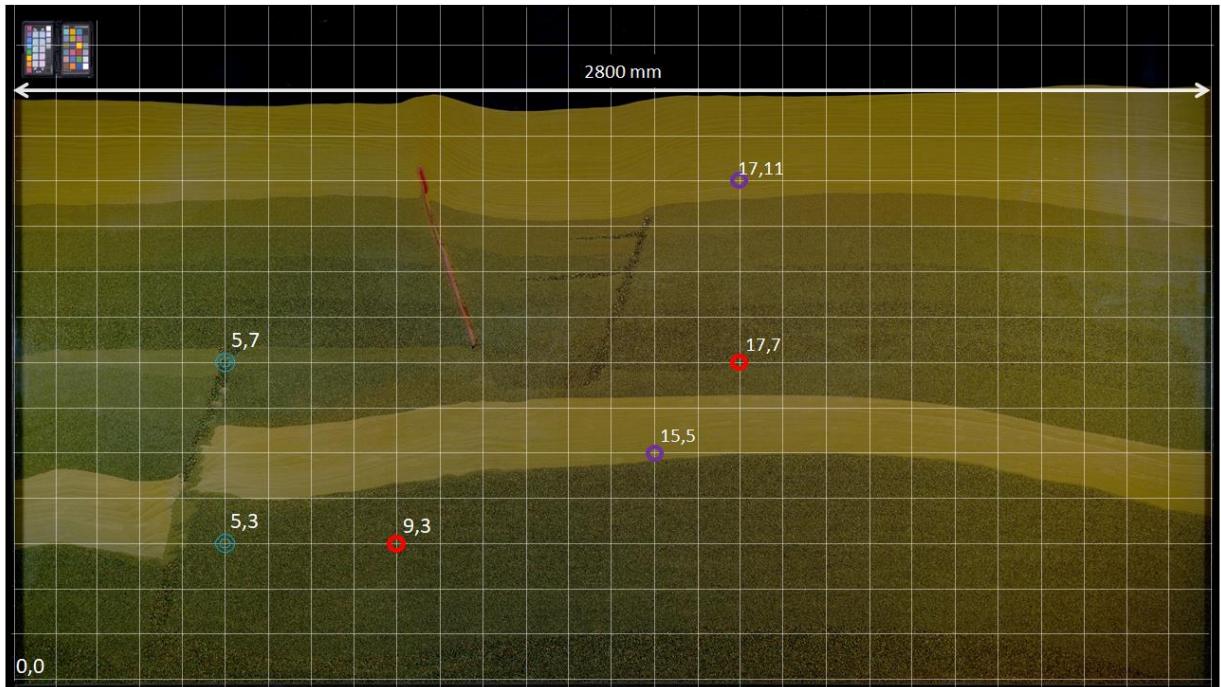


Figure 7: Port locations on the enhanced grid (each square is 100 x 100 mm). Horizontal length between far left and far right vertical lines is 2800 mm.

Note 1 *The visible length of the porous media is 2800 mm, but the active porous media extends 3 cm behind the black metal frame on each side.*

Note 2 *Shadows from the metal frame present on the left and right edge. No shadow at the bottom, but porous media extends 3 cm below bottom laser line.*

Note 3. *Port locations (circles) indicated. Color code: red is injection points, purple is pressure ports and open blue circles are additional well ports in the vicinity of the fault.*

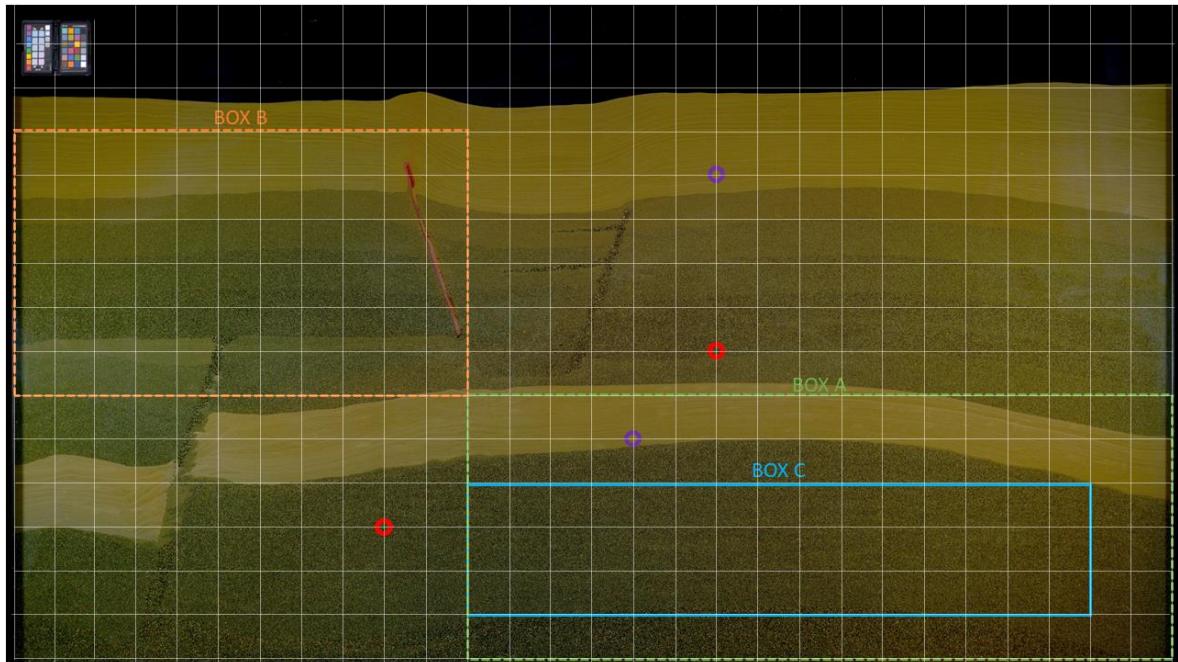


Figure 8: Box coordinates. Top right (TR), top left (TL), bottom right (BR), bottom left (BL)

Box A: TL (11,6) -> TR (28,6); BL (11,0) -> BR (28,0)

Box B: TL (0,12) -> TR (11, 12); BL (0,6) -> BR (11,6)

Box C: TL (11, 4) -> TR (26,4); BL (11,1) -> BR (26,1)

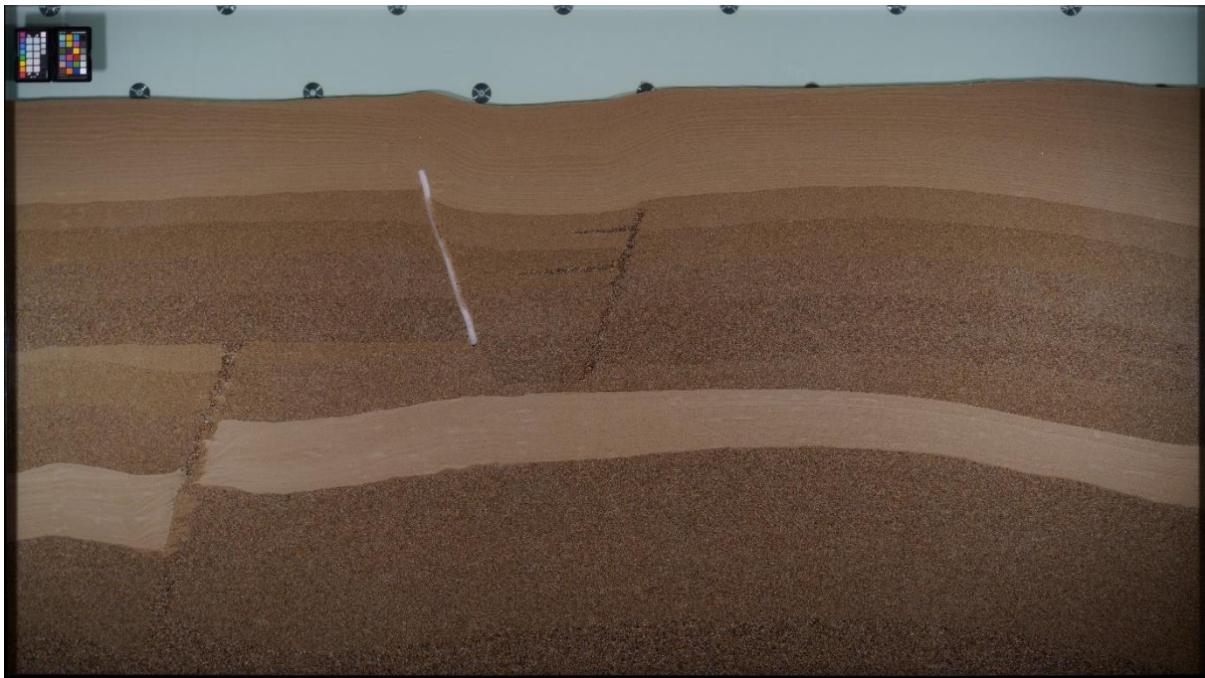
2.2.1 Sedimentation process

The dry, unconsolidated sand types were separately poured from the top into the water-filled (distilled water) void between the front and back panels. Each sand layer was built from the bottom and upwards, and features (predominantly change in dipping angle) were created by manipulating the layer during pouring, using guiding Styrofoam rectangles, funnels and plastic hoses. Mechanical manipulation (raking/scratching) was kept to a minimum and only in some areas in the vicinity of the faults. Faults (high permeable narrow bands of sand, at a steep angle + one sealed fault with silicone) were constructed through an iterative process, explained in more detail below.

The hydrostatic pressure during geometry assembly was 10 cm higher than at operating conditions. Hence, once the geometry was constructed, the water-level in the FluidFlower was lowered to operating water-level (kept constant during all injections). This led to an initial, pre-injection compaction of the sand. Further compaction was observed during subsequent days due to temperature fluctuations in the room where the rig is located, and during water (degassed distilled water) injection in all injection ports and technical ports in the bottom row.

2.2.2 Pre-injection flushing

Prior to CO₂ injection, the geometry was flushed with different aqueous fluids to further clean the sand and prepare for CO₂ injection experiments. The injection rates used (and associated pressures) were higher than during CO₂ injections. Hence, the geometry has experienced multiple flushing sequences, to satisfy compaction. Possible further compaction due to daily temperature fluctuations may occur, and will be documented. Such compaction is documented in the Figure 9 below.



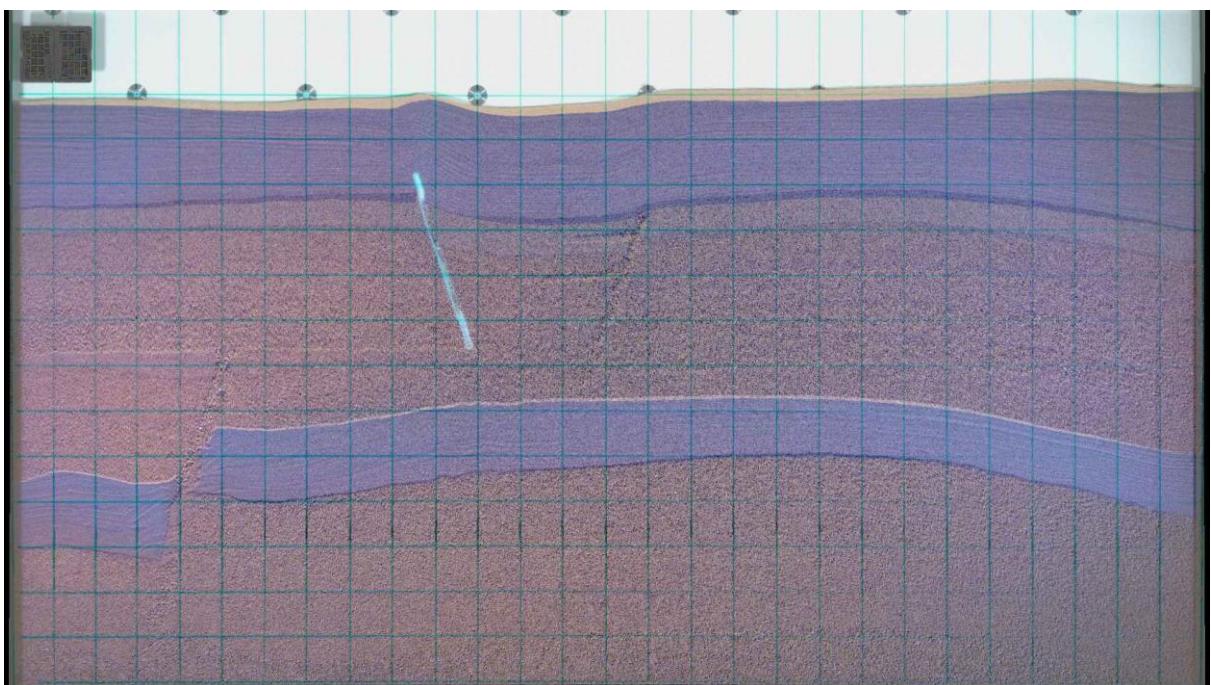
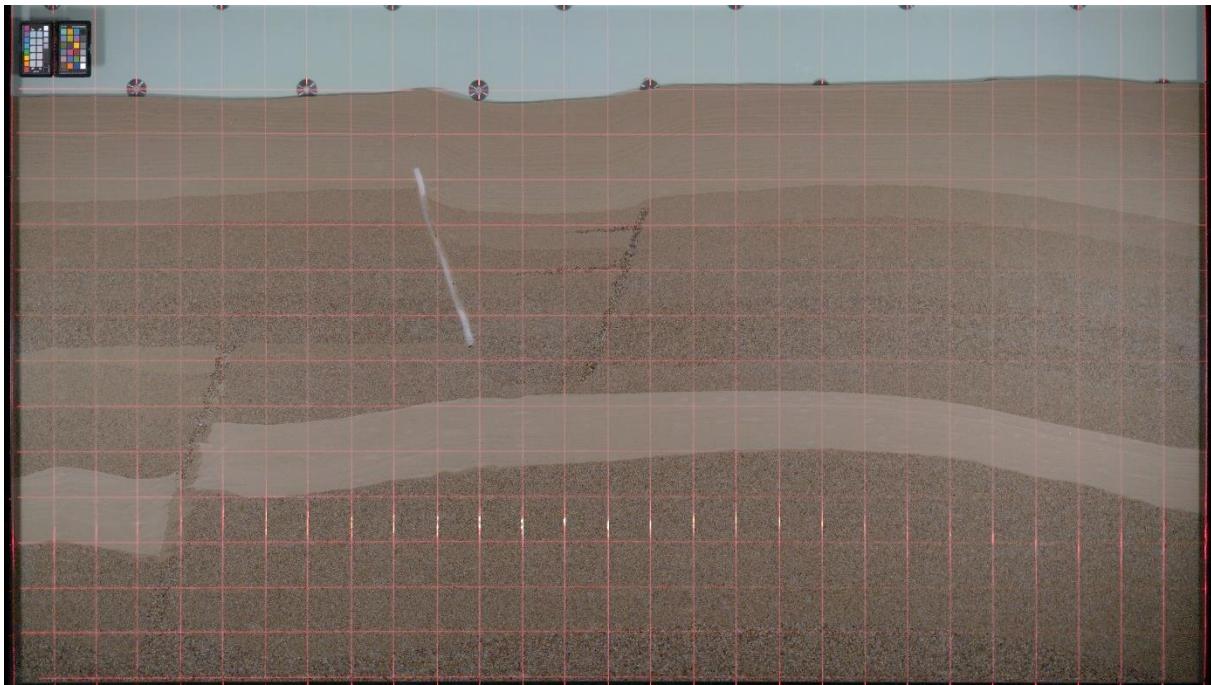


Figure 9: Fluidflower geometry (top) with laser grid (bottom) issued in September 2021. The differences compared with Figures 5-8 are due to further compaction after several flushing (aqueous) sequences, and temperature and pressure fluctuations in the laboratory.

All subsequent flushing sequences will be monitored with time-lapsed imaging to detect further, if any, compaction and reported.

2.3 Facies properties

The data for all sand types are reported in detail in Appendix A. Below follows a general description.

Sand labeled “Danish quartz sand” was purchased (totally 3.5 tons) from a commercial Norwegian supplier. Sand types used in the benchmark study were sieved from the supplied sand stock by the experimental team in Bergen - then washed (tap water) and acid washed until carbonate impurities were removed (time for each acid wash varied between 24 h to 72 h for the different sand types). The sands were then dried at 60 °C for at least 24 hours, and stored in 15 L plastic containers (washed and dried) until use. The different sand grain distributions are detailed in Table 1, along with sieve sizes, classification scheme. More information of each sand type in Appendix A.

Table 1: Sand type information.

Sand ID	Sieved grain size range [mm]	Grade	Class
ESF	*	Fine	Sand
C	0.5 - 0.71	Coarse (lower)	
D	0.71 - 1.0	Coarse (upper)	
E	1.0 - 1.41	Very coarse (lower)	
F	1.41 - 2.0	Very coarse (upper)	
G	2.0 - 2.8	Granules	Gravel

* ESF sand was not sieved (only washed, acid washed and dried) to maintain fine particles. Supplier provided 0.13 – 0.36 mm range; this was not accurate and grain sizes below and above cited range was observed.

2.3.1 Grain size distributions

The grain size distribution for each sand was analyzed after sieving (see Table 2 and Figure 10). The results demonstrate that each sand type had a wider grain size distribution than anticipated from the sieving process.

Table 2: Average smallest length, standard deviation and other key sand grain properties

	Average[mm]	Std [mm]	Max [mm]	Min [mm]	# grains analyzed	Area [mm ²]	Circularity (circle = 1.0)
Sand ESF*	0.20	0.11	0.47	0.06	100	-	-
Sand C	0.66	0.09	0.93	0.34	359	0.43	0.73
Sand D	1.05	0.14	1.53	0.68	474	1.11	0.68
Sand E	1.45	0.19	1.94	0.59	468	2.08	0.67
Sand F	1.77	0.31	2.56	0.54	284	3.05	0.64
Sand G	2.51	0.63	3.61	0.51	168	6.34	0.58

* Note that sand ESF was only measured manually (see Appendix A), focusing on the larger grains due to difficulties measuring small grains as a result of uncontrollable movement by static electricity.

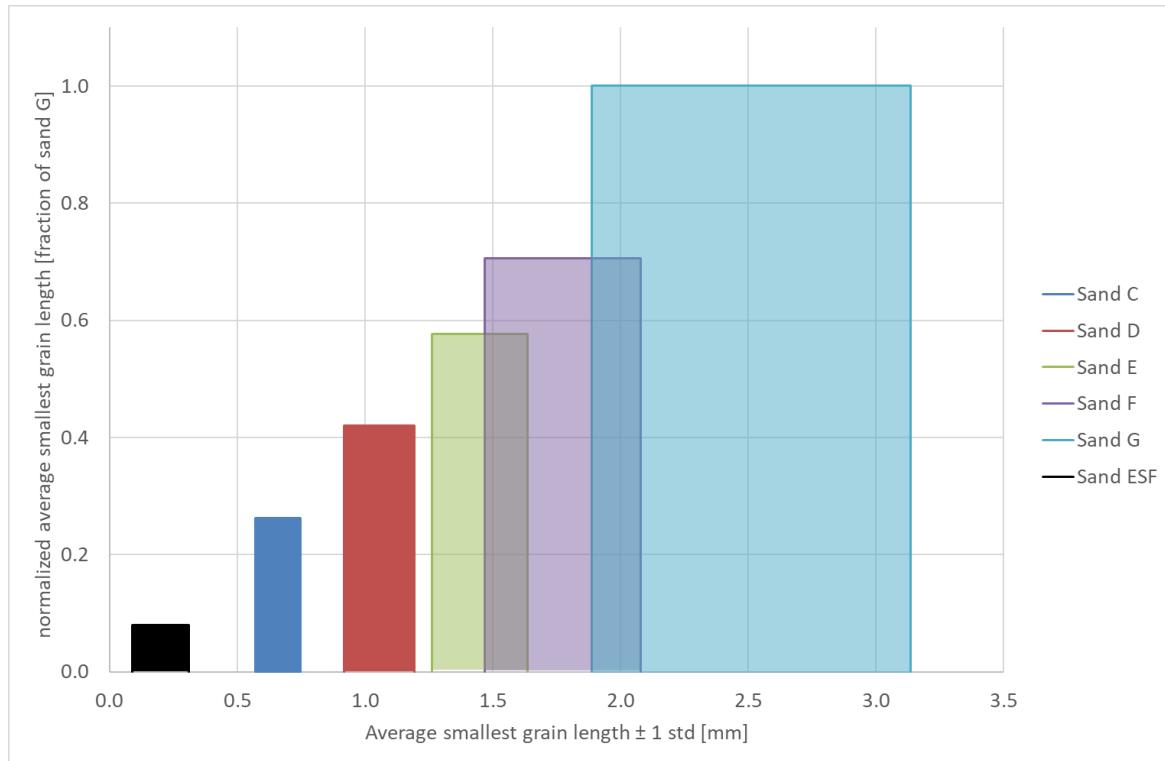


Figure 10: Distribution of smallest sand grain lengths of sands ESF, C, D, E, F and G plotted as average value \pm one standard deviation.

2.3.2 Absolute permeability measurements

The absolute permeability (see Table 3) of each sand type listed in Table 1 was measured prior to filling the FluidFlower rig. Measurements were performed using a glass cylinder (nominal dimensions: length 450 mm, diameter 20 mm). Each sand type was poured into a vertically positioned glass tube filled with degassed DI water, with bottom endpiece attached. Care was taken to not assert force to the sand when attaching the top endpiece, to avoid compaction beyond what would be expected in the FluidFlower rig. The glass tube was then placed horizontally and flushed with degassed DI water at a high rate (650 ml/h for 3 hours) to remove potential trapped air bubbles during assembly. Two absolute pressure transducers were placed at inlet and outlet, in addition to a differential pressure transducer. A fourth pressure transducer recorded changes in atmospheric pressure and temperature fluctuations. The permeability was measured in one direction, using 7 cycles of 5 ascending and descending constant volumetric injection rates (rates used was between 200 – 600 ml/h, with 100 ml/h increments) for 10 min (stable differential pressures were achieved). Between each cycle, water was injected with a constant rate of 200 ml/h for 10 min. The complete injection cycle was 13 hours. The flow rate was then reversed and the 7 cycles repeated.

Table 3: Measured absolute permeability of each sand (ESF, C, D, E, F, G)

	Average K [D]	Std [D]	Max [D]	Min [D]	# of cycles
Sand ESF	39	0.3	39	38	7
Sand C	293	8	302	278	7
Sand D	424	4	427	417	5
Sand E	708	39	799	682	7
Sand F	258	5	264	251	7
Sand G	488	33	568	466	7

Note that the reported absolute permeability only includes measurements performed vertically. Also note that sands E and G diverge from the theoretically expected trend ($K/D^2 = 6.8 \cdot 10^2 \text{ D/mm}^2$, as established for idealized random bead packings⁶) when plotting absolute permeability vs grain size (see Figure 11). From Table 2 we observe that the average smallest grain sizes for E and G are larger than the other sands, and we anticipate higher absolute permeabilities than the reported values. When comparing tracer transport (see Appendix B), we observe that the tracer (aqueous and fully miscible with reservoir water) advanced faster in sand G compared with sand F, and faster in sand F compared with sand E. Hence, the observed flow patterns in the benchmark geometry corroborate our expectation that the highest absolute permeability belongs to sand G (followed by F, E, D, C and ESF), in contrast to reported measurement of sands F and G. At this point we can only speculate on the reason for this discrepancy (e.g. wider sand grain distribution for E and G compared with the other sands; wall effects in glass column, experimental design, too low absolute pressures for accurate measurements). We will continue to measurement absolute permeability for all sands with the ambition to report true absolute permeabilities of all sands during the benchmark study.

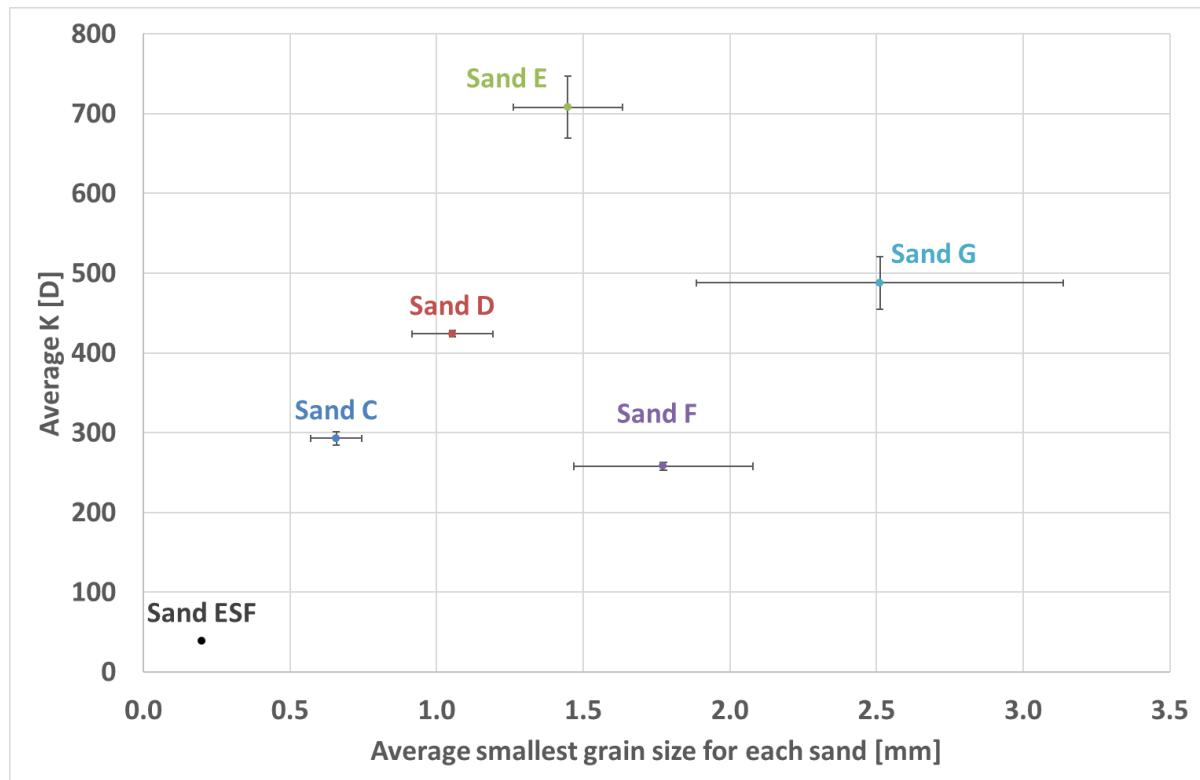


Figure 11: Distribution of measured absolute permeability (Table 3) plotted against average smallest grain size (Table 2), with \pm one standard deviation. Note that sands F and G do not follow the trend, and these values are not considered reliable.

⁶ Bosl, Dvorkin, Nur, «A study of porosity and permeability using a lattice Boltzmann method», GRL 1998.

2.3.3 Porosity measurements

Porosity was calculated using the following procedure: the weight of a water-filled (without sand) glass column, with a predetermined water height corresponding to the end of the top endpiece, was recorded. The water height was kept constant when sand was poured into the glass column; i.e., water was removed from the column with a syringe when sand was added. The sand was filled up to the predetermined level (the same as the initial water height). The volume of the water removed was recorded and equals the system pore volume. The porosity (see Table 4) is calculated from the ratio of the pore and bulk volume (calculated from weight and measured column dimensions). Average permeability for each sand type with standard deviation will be reported, along with measured porosity values.

Table 4: Calculated porosity values for each sand.

	Porosity	# of measurements
Sand ESF	0.44	1
Sand C	0.43	1
Sand D	0.44	1
Sand E	0.45	1
Sand F	0.43	1
Sand G	0.46	1

2.3.4 Unsteady-state relative permeability measurements - endpoints

The unsteady-state end-point relative permeabilities (Table 5) during both drainage and imbibition was using CO₂-saturated water and CO₂. The average end-point fluid saturation was measured using from volumetric and weight measurements. The following procedure was followed: The glass tube (identical as above) was placed vertically, with the bottom end piece attached. The tube was filled with degassed DI water and the sand was poured in from the top and the top end piece attached, similar to absolute permeability measurements described above. Degassed DI water was then injected from the top with a high rate (650 ml/h for 3 hours) to remove potential gas bubbles. Absolute permeability was measured in a vertical position, with degassed DI water injected from the top using several cycles, as described above. When the cycles of absolute permeability were finished, the degassed brine was replaced in the sand column with CO₂-saturated DI water; gaseous CO₂ was injected into the DI water bottle, and the pH was measured in the bottle to monitor CO₂ saturation levels. The CO₂-saturated DI water was then injected from the top to miscibly displace the degassed DI water in the column. Outlet tube(s) were placed at the same height as the inlet, and it was confirmed that water was not produced from the column due to gravity before CO₂ was injected. The weight of the water-saturated sand column was recorded.

End-point drainage: CO₂ was injected from the top with a constant volumetric (mass) rate. The absolute inlet and outlet pressure, in addition to the differential pressure, were recorded during drainage. Water production was monitored by weight during injection, and CO₂ was injected until no further water production was recorded. Some tests also include rate bumps (increased injection rates after gas breakthrough), with no additional water production recorded. The weight of the partially water saturated sand column was recorded.

End-point imbibition: After the drainage process was completed, the injection was switched to constant volumetric rate (600ml/h) of CO₂-saturated water from the top. The pressures and weight of sand column were recorded during injection. When no additional gas was produced, the final sand column weight was recorded and an injection cycle to measure water endpoint permeability was conducted. Once completed, the weight of the sand column was identical ($\pm 0.1\text{g}$) to previous weight.

Table 5: Calculated endpoint relative permeabilities for gas and water for each sand.

	K [D] (from Table 3)	# of measurements	Endpoint gas (CO ₂)		Endpoint water	
			Swi	k _{rel} gas	1-Sg	k _{rel} water
Sand ESF	39	1	-	-	-	-
Sand C	296	1	0.15	0.20	-	-
Sand D	431	1	0.17	0.20	-	-
Sand E	708	1	0.21	0.10	-	-
Sand F	259	1	0.03	0.11	-	-
Sand G	495	1	0.14	0.16	0.9	0.72

Note that there are large uncertainties associated with the measurements of endpoint relative permeabilities reported in Table 5. We are already in the process of repeating these measurements, and to identify why the measured K and k_{rel} are not as expected. For sand F there are especially large uncertainties due to suboptimal procedures (these were corrected for subsequent measurements, we measured sand F first). There are also missing data: all measurements for sand ESF (due to sand column collapse) and end point values for 1-Sg and k_{rel} water for all sands except sand G. A complete and updated Table 5 will be made available on Github when ready.

Capillary entry pressures

The capillary entry pressure to gas (see Table 6) was measured for each sand type, based on gas column break-through experiments in a down-scaled FluidFlower under equivalent sedimentary protocol as for the FluidFlower experiment. A range of geometries were investigated, from steep anticlines to flat anticlines (as in Figure 12). The porous media was filled with pH-sensitive dye (blue-green color) and CO₂ flow was monitored with time-lapsed imaging to detect the maximum gas column height under the sealing layer (ESF sand, see Table 1) and between sand types. The observed gas column height (in m) was converted to pressure to provide the entry pressure for each sand type.

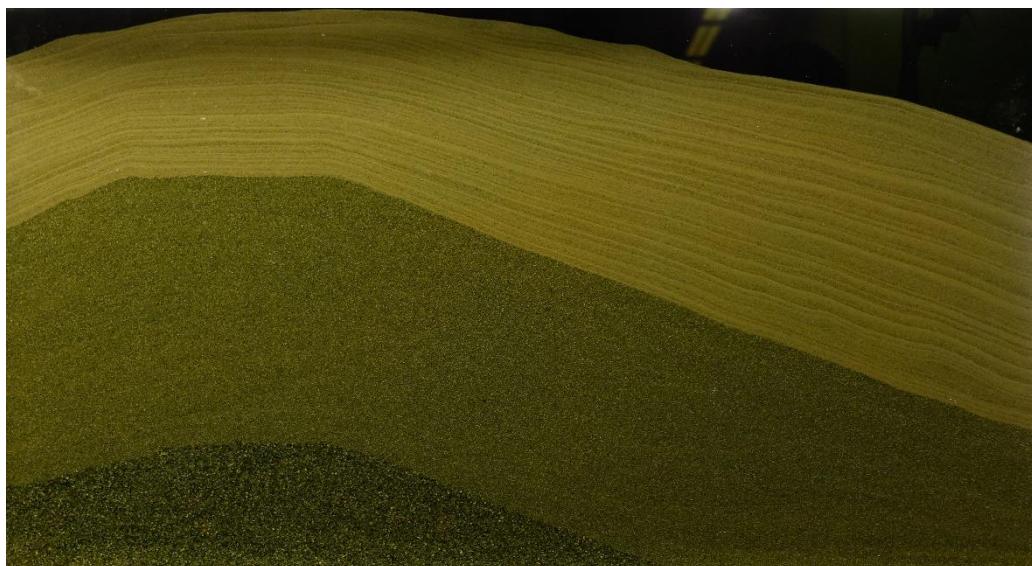


Figure 12: Image of one of the geometries used to observe and calculate capillary entry pressure for the different sands. Note that the image shows 100% filled with aqueous phase, without CO₂ gas.

Table 6: Calculated capillary entry pressure for each sand type.

	Measured gas column height [m]	P _c entry pressure [N/m ²]	P _c entry pressure [millibar]
Sand ESF	0.015	1471.5	15
Sand C	0.03	294.3	3
Sand D	0.01	98.1	1
Sand E	0	-	-
Sand F	0	-	-
Sand G	0	-	-

Note 1

Additional measurements are ongoing to provide the entry pressure values with standard deviation.

2.4 Fault Facies

The upper, high permeable fault (see Figure 4) was created with sand G (Table 1); the upper left fault is sealing (with silicone, see Figure 13) and the lower left fault is characterized as heterogeneous. The faults (expect the sealing one) were created by using Styrofoam shapes, funnels and plastic hoses as mentioned above. Fault sand(s) were filled simultaneously as the adjacent layers, with the sand layer within the fault always above (in the order of 5 cm) the adjacent layer. The Styrofoam shapes were lifted step-wise during filling, and sand in the adjacent layer moved next to the fault sand.



Figure 13: Close up of “sealed fault”. Note that after sand compaction (due to temperature fluctuations in the laboratory and expanding front/back panels – in the millimeter range) the fault is no longer sealing along its entire length. It is indicated (white arrow) where sand grains touch each other from both sides of the fault. The fault appears sealing in remaining parts.

2.5 Fluid properties

The following fluids are used:

- Water composition: distilled water + pH sensitive salts (density ~1002 kg/m³)
- Gaseous CO₂: high purity (99.999% - 5.0 purity)

All experiments are conducted at room temperature (fluctuating, will be measured and reported, aim for approximately 20°C) and atmospheric pressure at the free water table at the top of the experiment. We will attempt to keep temperature fluctuations limited, however due to the available laboratory space, a constant temperature cannot be guaranteed. Variations in atmospheric pressure and temperature will be supplied to the benchmark participants at the completion of the experiments.

2.6 Operational conditions

The geometry is initially 100% water-filled. Gaseous CO₂ will be injected into the lower injection point (see Figure 3) with a constant mass flow (10 ml/min; unit standard milliliters per minute – sccm). After a delay, CO₂ will be injected into the upper injection point with a constant mass flow.

Tentative CO₂ injection protocol (*may be updated after experiment has been run*):

1. Start CO₂ injection in port (9,3) @ 10 ml/min and run for 24 hrs
2. Start CO₂ injection in port (17,7) @ 10 ml/min after a 12 hrs delay relative to start in (9,3), and run for 12 hrs.
3. Monitor CO₂ transport for 4 days

2.7 Well test data

To allow for calibration of the numerical models, and in line with real field operations, we provide 2 sets of calibration data. These are detailed in Appendix B.

2.7.1 Pressure response

Well-test data associated with active injection wells (the two CO₂ injection points) used in this benchmark are reported. For each of the well tests, one well will be rate-controlled, and pressure data will be measured at the other injection point, as well as the two pressure ports in the seal (ports 15, 5 and 17, 11, cf. Figure 7) and the two fault-related ports (ports 5,3 and 5,7, cf. Figure 7). The well-tests were run as square-pulses repeated three times in each injection point. Pressure transducers were mounted on each port described above, sometimes with two transducers in the same point to assess uncertainty. The pressure transducers were run in gauge mode, and zeroed before injection start. Location, pressure range and uncertainty for each transducer will be provided, and raw, unfiltered pressure data will be supplied as a text-file with the following data-structure:

Time-stamp	Injection point ID	Rate (injector)	Press 1	Press 2	...	Pressure n
------------	--------------------	-----------------	---------	---------	-----	------------

2.7.2 Tracer flow

We supply one set of tracer flow data. For the tracer flow data, tracer (pH indicator, density ~1002 kg/m³) will be injected with a constant volumetric flow rate. Data is supplied as image files (Appendix B). A video of the injection is available at <https://github.com/fluidflower>

2.8 Measurables

The following measurables shall be reported by the experiment, and predicted by the benchmark participants, details of which, including time resolution, are described in Section 3.

2.8.1 Pressure

Pressure shall be reported at each of the two pressure sensors in [N/m²].

2.8.2 Phase composition

The distribution of the CO₂-phase shall be reported in boxes labeled A and B in Figure 4. The phase distribution, in SI units (unitless for volume fractions, kg for mass quantitites), shall be reported in the following categories: Mobile free phase (CO₂ at saturations for which the relative permeability exceeds 0), immobile free phase (CO₂ at saturations for which the relative permeability equals 0), dissolved (CO₂ in water phase) and seal (CO₂ in any form in yellow sand regions in Figure 4). The sum of the three first categories, mobile, immobile and dissolved, shall equal the total mass of CO₂ in the box.

2.8.3 Convection

For the box labeled Box C in Figure 3, the convective mixing shall be reported as the integral of the magnitude of the gradient in relative concentration of dissolved CO₂. In other words, if mass-fraction of CO₂ in water is denoted χ_c^w , and the dissolution limit is denoted $\chi_{c,\max}^w$, then the following quantity M shall be reported

$$M(t) \equiv \int_C \left| \nabla \left(\frac{\chi_c^w}{\chi_{c,\max}^w} \right) \right| dx$$

3. Data reporting and interaction

All result data will be uploaded by the participants to Git repositories within the GitHub organization <https://github.com/fluidflower>. Each participant will get write access to a dedicated repository named after his/her institution, e.g., <https://github.com/fluidflower/stuttgart>. During the blind phase, only the participants themselves will have access to their respective repositories. For the synchronization phase, read access to all participant repositories will be granted for all participants. After the workshop in February, the repositories will be further opened to include also the results from the physical experiments. Upon submission of the papers, the relevant repositories will be turned public.

The reported data will be analyzed in two respects: Both in terms of an intercomparison of numerical simulation capability, but also in terms of our ability to correctly assess key properties of the system. Consequently, we establish both a dense and a sparse reporting protocol.

3.1 Dense data – basis for numerical intercomparison study

The reported dense data should be the best numerical data available, in the sense of finest grid resolution, most representative physics, and so forth.

All measurables identified in section 2.8 shall be reported at 10-minute intervals starting at the initial injection and lasting 120 hours. The data is expected in csv format in the repository in a file `time_series.csv` of the form

```
# t, p_1, p_2, mob_A, imm_A, diss_A, seal_A, <same for B>, M_C  
0.000e+00, 1.234e+56, 1.234e+56, <...>  
6.000e+02, 1.234e+56, 1.234e+56, <...>  
...
```

according to the measurables defined in section 2.8.

Additionally, a map of the phase compositions shall be reported for each 24 hours after injection starts. While the actually employed computational grids are generated by the participants individually, these maps shall be reported on a uniform Cartesian grid of 286 by 123 cells (1 cm by 1 cm grid cells from the bottom of the domain) in order to facilitate comparisons. For each temporal snapshot indicated by X hours, X = 24, 48, ..., cell values should be provided in csv format in a file `spatial_map_<X>.h.csv` of the form

```
# x, y, gas saturation [-], CO2 concentration in water [kg/m3]  
5.000e-03, 5.000e-03, 1.234e+56, 1.234e+56  
1.500e-02, 5.000e-03, 1.234e+56, 1.234e+56  
...  
2.855e+00, 5.000e-03, 1.234e+56, 1.234e+56  
5.000e-03, 1.500e-02, 1.234e+56, 1.234e+56  
1.500e-02, 1.500e-02, 1.234e+56, 1.234e+56  
... ...  
2.845e+00, 1.225e+00, 1.234e+56, 1.234e+56  
2.855e+00, 1.225e+00, 1.234e+56, 1.234e+56
```

The origin of the coordinate system should be located in the lower left corner with the x-axis positively oriented towards the right and the y-axis positively oriented towards the top. We note that the relevant laser grid elements depicted in the figures of the preceding sections extend from approximately (0.03, 0.03) to (2.83, 1.23) in this coordinate system.

3.2 Sparse data – basis for predictive capability assessments

The reported data should be given to the best ability of each research groups, and need not be direct simulation output. Indeed, it is permitted (encouraged) to use one's own experience to temper the numerical predictions, as if this was an analysis of a real field operation.

The following sparse data is requested:

1. *As a proxy for assessing risk of mechanical disturbance of the overburden:* Maximum pressure at sensor number 1 and 2.
2. *As a proxy for when leakage risk starts declining:* Time of maximum mobile free phase in Box A.
3. *As a proxy for our ability to accurately predict near well phase partitioning:* All quantities defined in Section 2.7.2 in Box A at 72 hours after injection starts.
4. *As a proxy for our ability to handle uncertain geological features:* All quantities defined in Section 2.7.2 in Box B at 72 hours after injection starts.
5. *As a proxy for our ability to capture onset of convective mixing:* Time for which the quantity M defined in Section 2.7.3 first exceeds 110% of the width of Box C.

6. *As a proxy for our ability to capture migration into low-permeable seals:* Total mass of CO₂ in the top seal facies (areas marked yellow in the sketch) at final time within Box A.

Each of the sparse data shall be reported as six numbers, representing the prediction of the mean quantity as obtained by the experiments (stated in terms of P10, P50 and P90 values), as well as the prediction in the standard deviation of the quantity over the ensemble of experiments (again stated as P10, P50, and P90 values). Loosely speaking the predictions of mean values assess the capability of predicting the various measurables, while the predictions of the standard deviations address the extent to which these quantities are deterministic.

As basis for generating the predictions and uncertainties, any preferred methodology may be chosen (ensemble runs, some methods from uncertainty quantification, human intuition from experience or any combination of these).

The quantities are expected to be uploaded to the repository in form of a csv file `sparse_data.csv` of the form

```
# idx, p10_mean, p50_mean, p90_mean, p10_dev, p50_dev, p90_dev
1a, 1.234e+56, <...>, # pressure at sensor 1 [N/m2]
1b, 1.234e+56, <...>, # pressure at sensor 2 [N/m2]
2, 1.234e+56, <...>, # time of max mobile free phase in Box A [s]
3a, 1.234e+56, <...>, # mobile free phase in Box A at 72h [kg]
3b, 1.234e+56, <...>, # immobile free phase in Box A at 72h [kg]
3c, 1.234e+56, <...>, # dissolved in water in Box A at 72h [kg]
3d, 1.234e+56, <...>, # seal in Box A at 72h [kg/m2]
4a-d <same for Box B>
5, 1.234e+56, <...>, # time when M exceeds 110% of Box C's width [s]
6, 1.234e+56, <...>, # total mass of CO2 in the top seal facies [kg]
```

3.3 Qualitative data – questionnaires

During the benchmark period, and in particular during the workshops, questionnaires will be provided to gather contextual information. These will address both issues of numerical modeling and simulation, as well as issues of choices and judgements made for assessing the confidence intervals for the sparse data.

3.4 Reporting deadlines

The blind data shall be uploaded to the respective repository by December 12, 2021. Revised data, if desired, can be submitted in the same format by March 10, 2022.

Each participating group will report on their blind data at the virtual workshop on December 15, 2021, and at the real-life workshop in the week March 14-18, 2022. The workshops will be structured so that our mutual judgement of the quality of both your own results, as well as those of your peers, can be assessed.

3.5 Limited interaction

In order to protect the integrity of the results, we ask for dedicated communication rules during the different phases of the benchmarking process. To facilitate remote communication between

participants, and also to store this communication for evaluating the benchmarking process, a Discord server has been set up at <https://discord.gg/8Q5fZS3T47>. Apart from a general channel that is initially open to everyone involved, a private channel is installed for each participant which should be used for communicating with the benchmark organizers.

For the preparation phase before the kick-off meeting, no restrictions apply. This phase should be used by all participants and the experimental group to shape this benchmark description. During the blind phase, each participating group acts on its own and we ask that no communication on the subject takes place between different participating groups and with the experimental group. For the participants, the general Discord channel will be read-only. Any questions of clarification should be directed to Bernd Flemisch by the respective private Discord channel, who will post the replies on the general channel if they are of common interest. After the virtual workshop in November, the synchronization phase starts, during which the participating benchmarking groups can communicate freely, preferably over the Discord channels for monitoring and archival of the process. Still, no communication with the experimental group will be allowed before the real workshop in February.

Appendix A: Facies descriptions

This appendix gives detailed information of the six sand types used in the geologic setup. Analysis and further measurements are ongoing as this description (version 2.3) is issued.

Facies 1 – Sand F (Main Reservoir)

Grain size distribution

- Sieve sizes used: 1.41 mm – 2.0 mm

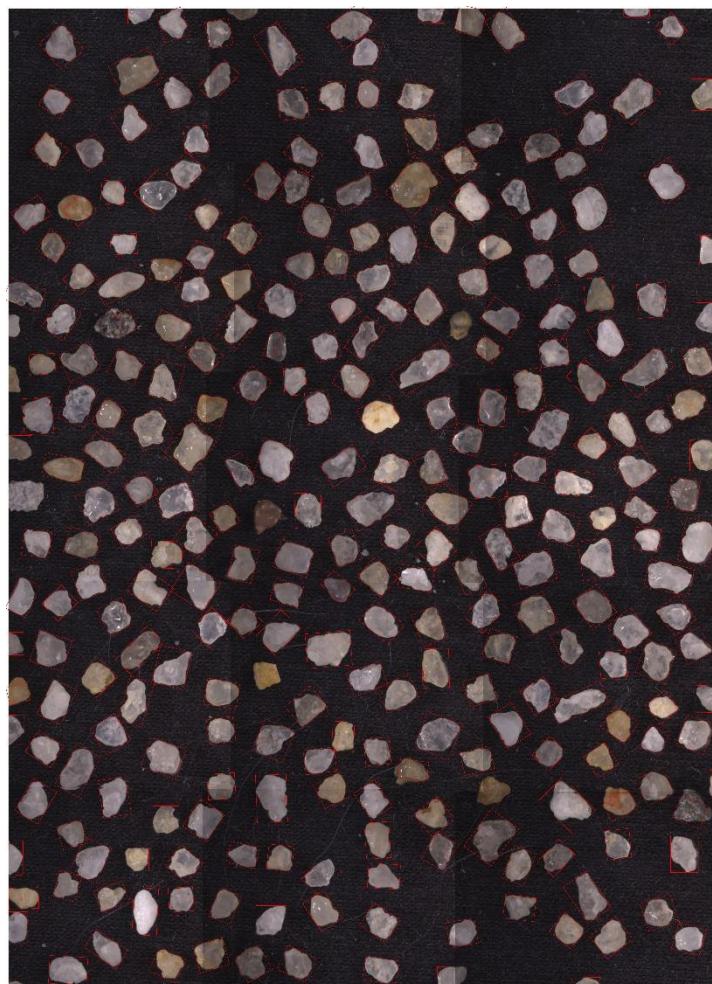


Figure A1 - A microscope image of **Sand F** and measured values for each of the 284 sand grains. The image shows the grains bounded by rectangles. To consider the rotation of the grains, the rectangles are drawn with minimum area. Grains that have similar color (dark grey and black) to the background on a binary image were not segmented. Grains with a partially similar color to the background was only partially segmented so that the calculated lengths for these grains are less than the actual length. Grains located close to the border, only partially visible in the image, are included in the analysis and contribute to the population of very small lengths. This will be corrected in future analysis.

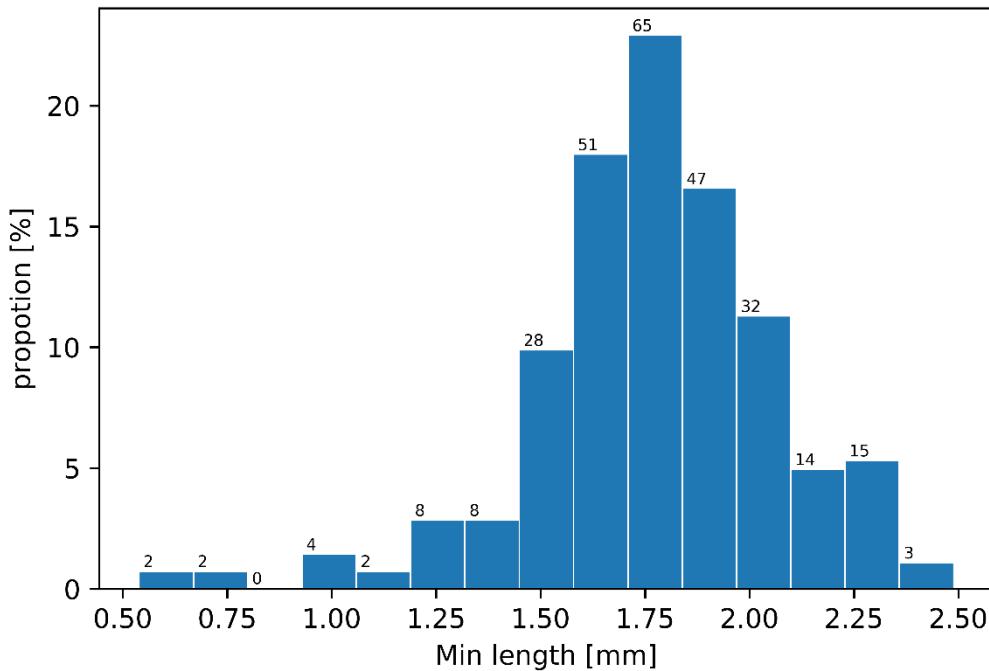
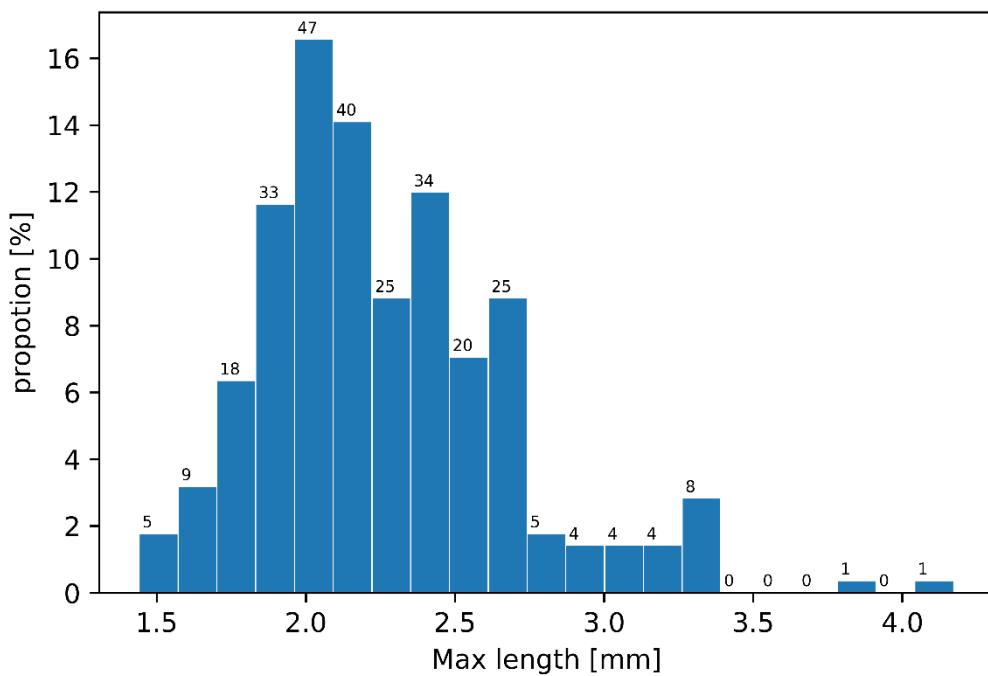


Figure A2 - Histograms for long (top) and short (bottom) grain lengths measured for **Sand F**. Number

of grains in each population is indicated.

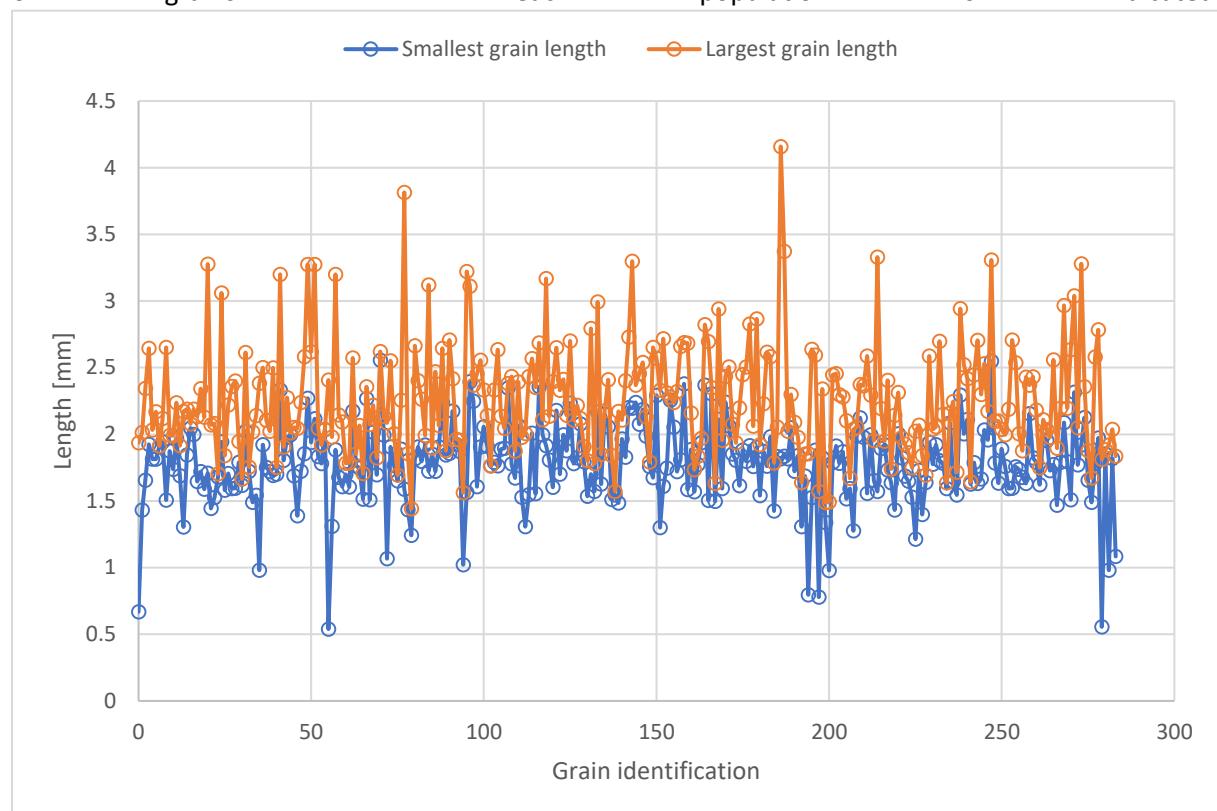


Figure A3 – Graphs for long (top, blue) and short (bottom, gray) grain diameters measured for **Sand F** using image in Figure A4. Each graph has 284 measurements.

Facies 2 – Sand E

Grain size distribution

- Sieve sizes used: 1.0 mm – 1.41 mm



Figure A4 - A microscope image of **Sand E** and measured values for each of the 468 sand grains. The image shows the grains bounded by rectangles. To consider the rotation of the grains, the rectangles are drawn with minimum area. Grains that have similar color (dark grey and black) to the background on a binary image were not segmented. Grains with a partially similar color to the background was only partially segmented so that the calculated lengths for these grains are less than the actual length. Grains located close to the border, only partially visible in the image, are included in the analysis and contribute to the population of very small lengths. This will be corrected in future analysis.

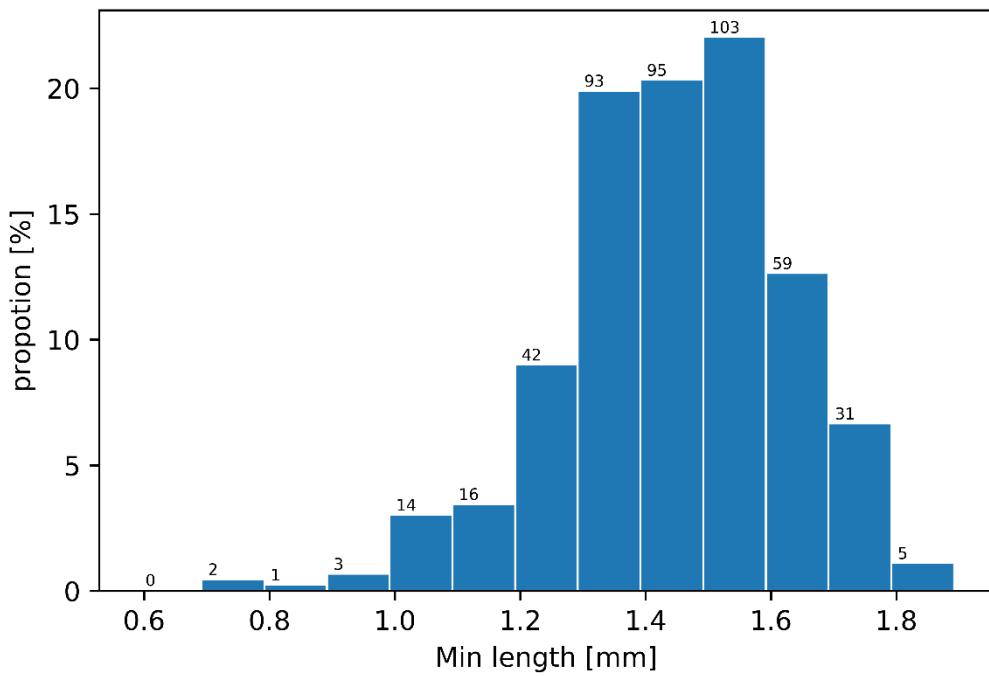
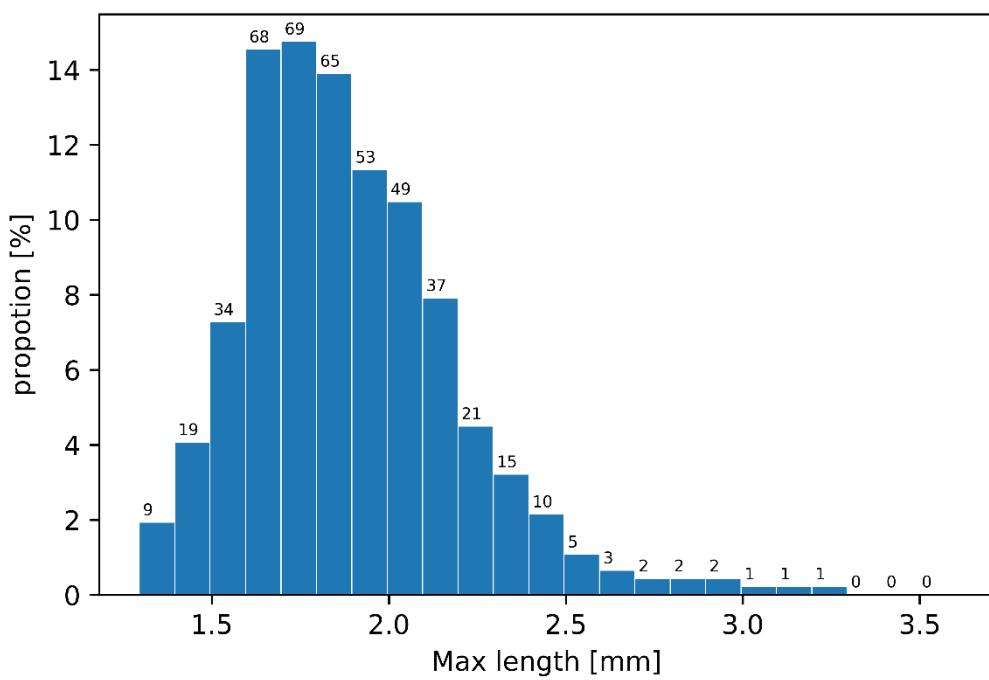


Figure A5 - Histograms for long (top) and short (bottom) grain lengths measured for **Sand E**. Number of grains in each population is indicated.

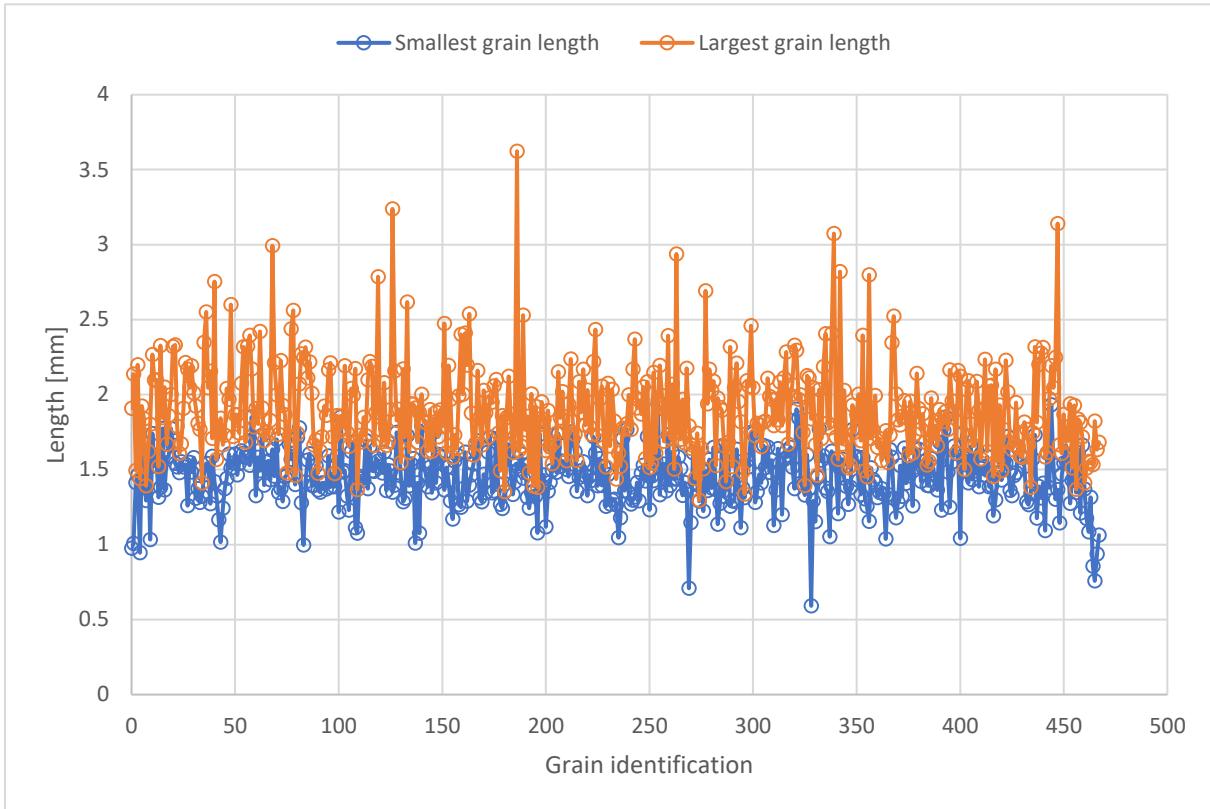


Figure A6 – Graphs for long (top, blue) and short (bottom, gray) grain diameters measured for **Sand E** using image in Figure A4. Each graph has 468 measurements.

Facies 3 – Sand D

Grain size distribution

- Sieve sizes used: 0.71 mm – 1.00 mm



Figure A7 - A microscope image of **Sand D** and measured values for each of the 474 sand grains. The image shows the grains bounded by rectangles. To consider the rotation of the grains, the rectangles are drawn with minimum area. Grains that have similar color (dark grey and black) to the background on a binary image were not segmented. Grains with a partially similar color to the background was only partially segmented so that the calculated lengths for these grains are less than the actual length. Grains located close to the border, only partially visible in the image, are included in the analysis and contribute to the population of very small lengths. This will be corrected in future analysis.

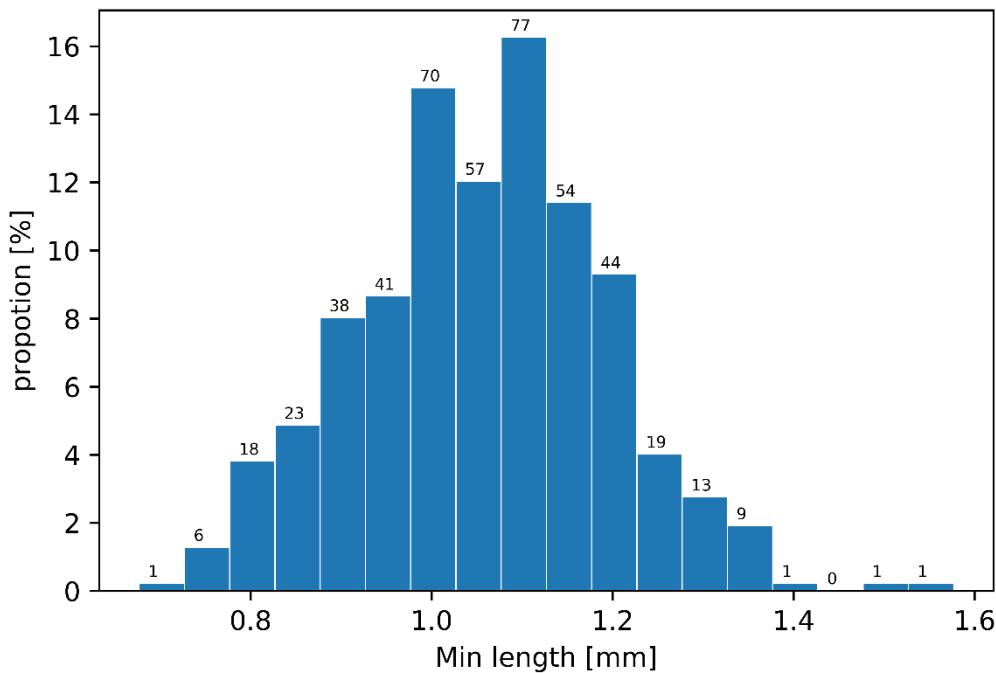
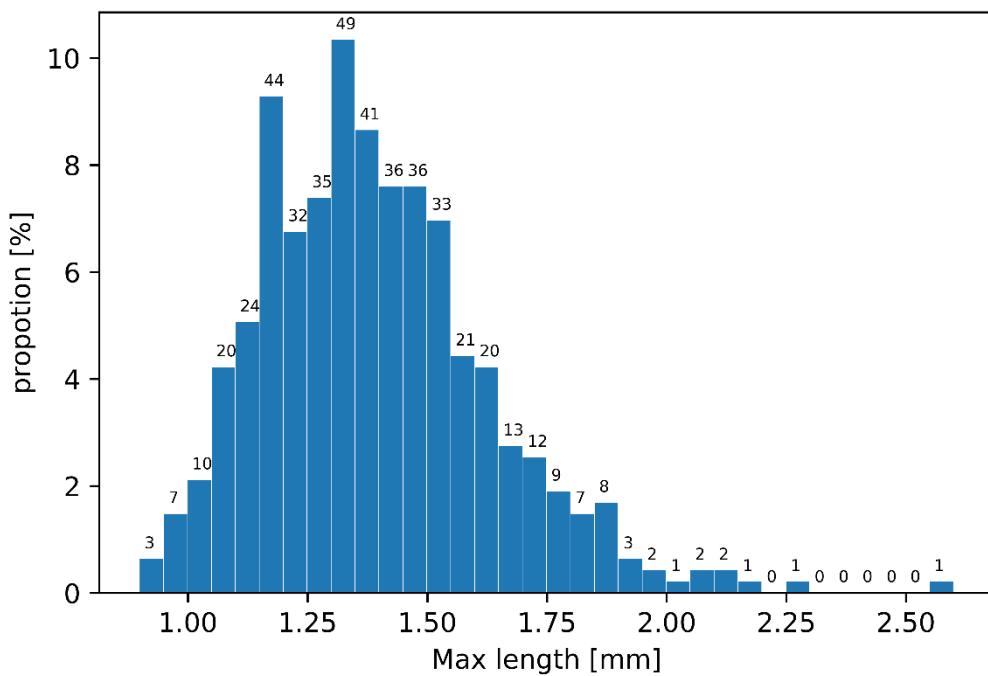


Figure A8 - Histograms for long (top) and short (bottom) grain lengths measured for **Sand D**. Number of grains in each population is indicated.

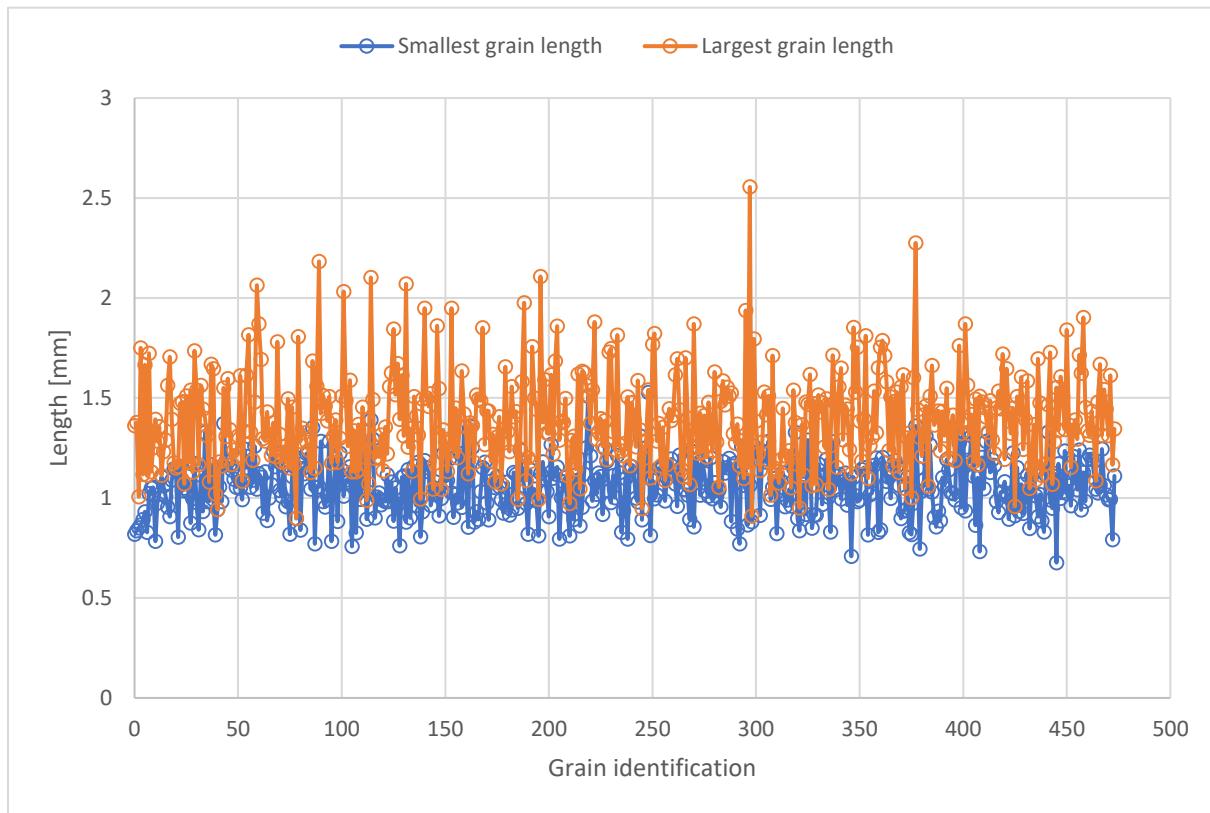


Figure A9 – Graphs for long (top, blue) and short (bottom, gray) grain diameters measured for **Sand D** using image in Figure A4. Each graph has 474 measurements.

Facies 4 – Sand C

Grain size distribution

Sieve sizes used: 0.50 mm – 0.71 mm

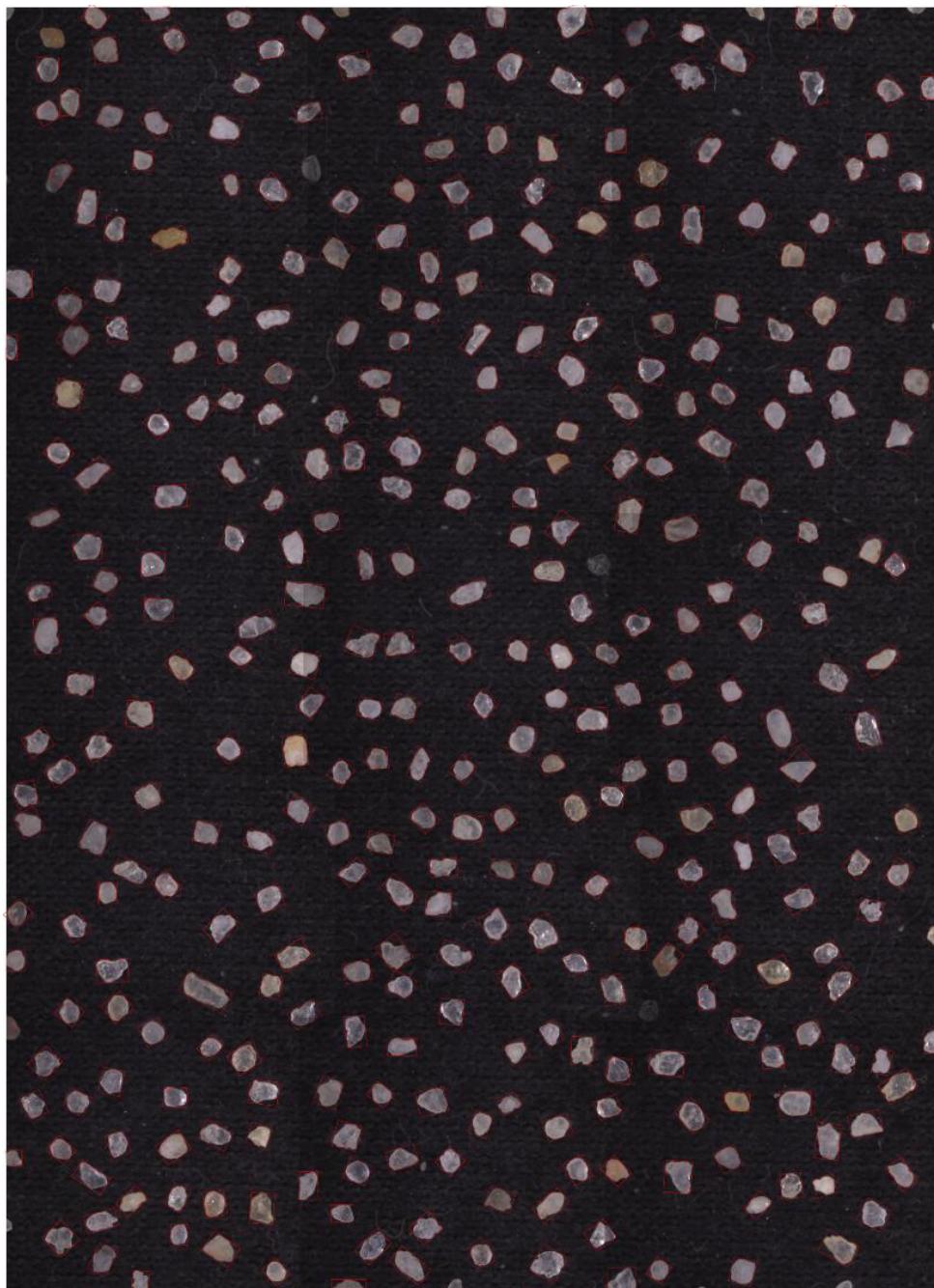


Figure A10 - A microscope image of **Sand C** and measured values for each of the 359 sand grains. The image shows the grains bounded by rectangles. To consider the rotation of the grains, the rectangles are drawn with minimum area. Grains that have similar color (dark grey and black) to the background on a binary image were not segmented. Grains with a partially similar color to the background was only partially segmented so that the calculated lengths for these grains are less than the actual length. Grains located close to the border, only partially visible in the image, are included in the analysis and contribute to the population of very small lengths. This will be corrected in future analysis.

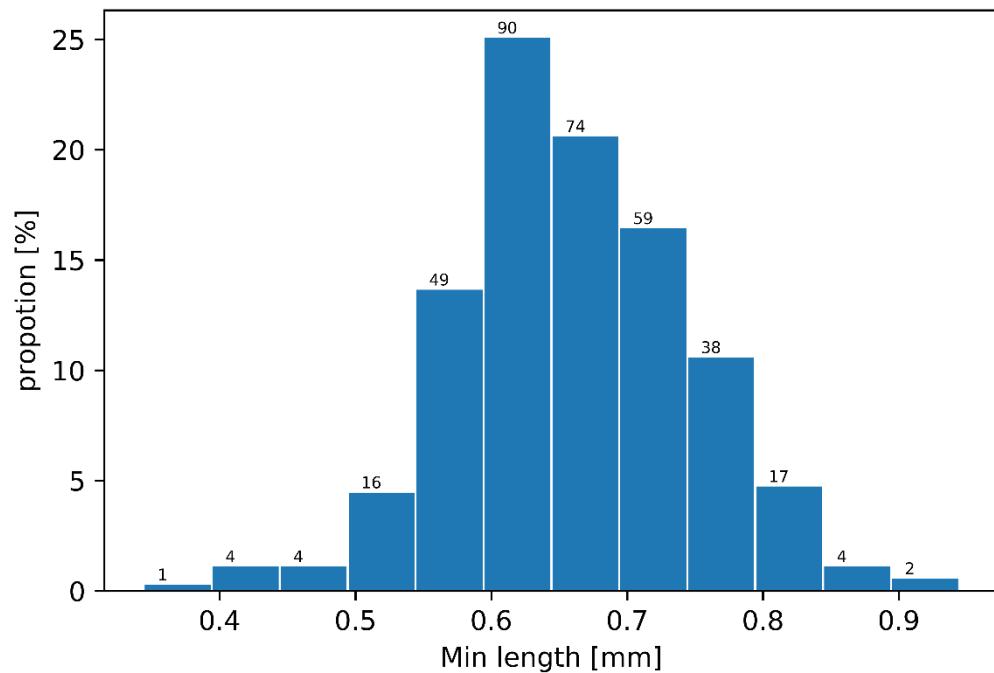
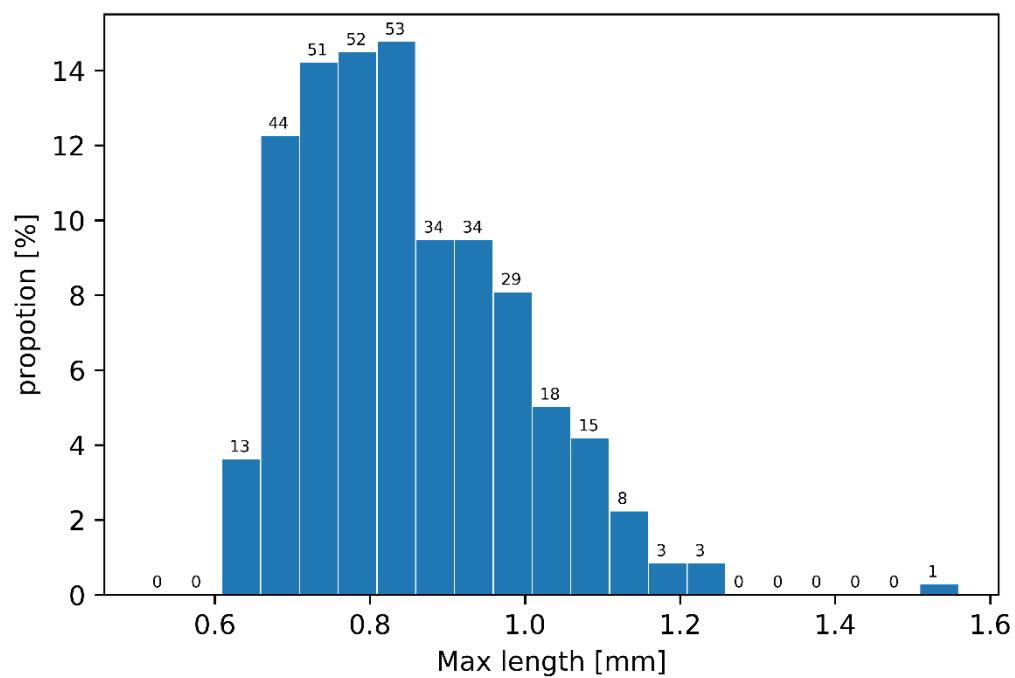


Figure A11 - Histograms for long (top) and short (bottom) grain lengths measured for **Sand C**. Number of grains in each population is indicated.

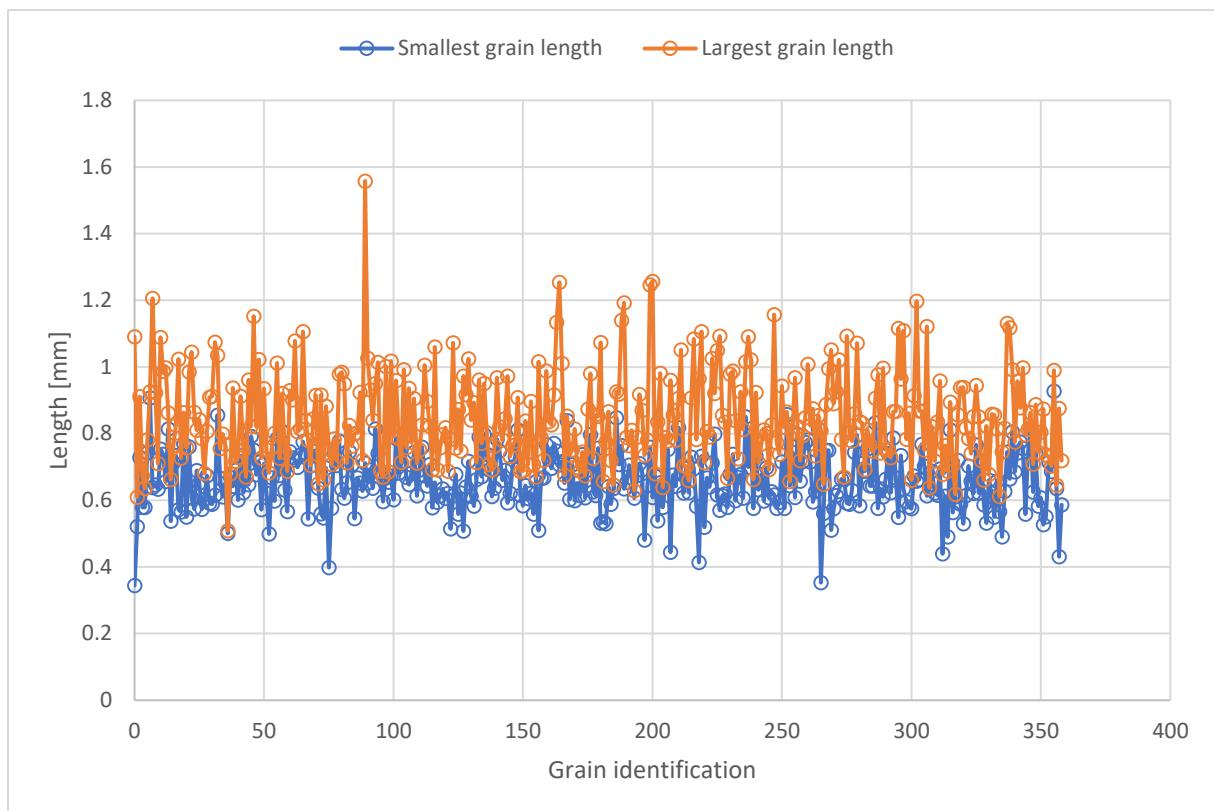


Figure A12 – Graphs for long (top, blue) and short (bottom, gray) grain diameters measured for **Sand C** using image in Figure A4. Each graph has 359 measurements.

Facies 5 – Sand ESF

Grain size distribution



Figure A13 - A microscope image of **Sand ESF** and measured values for each of the 100 sand grains. Because the sand grains are not perfect spheres, two lengths were measured and labelled “long” and “short”.

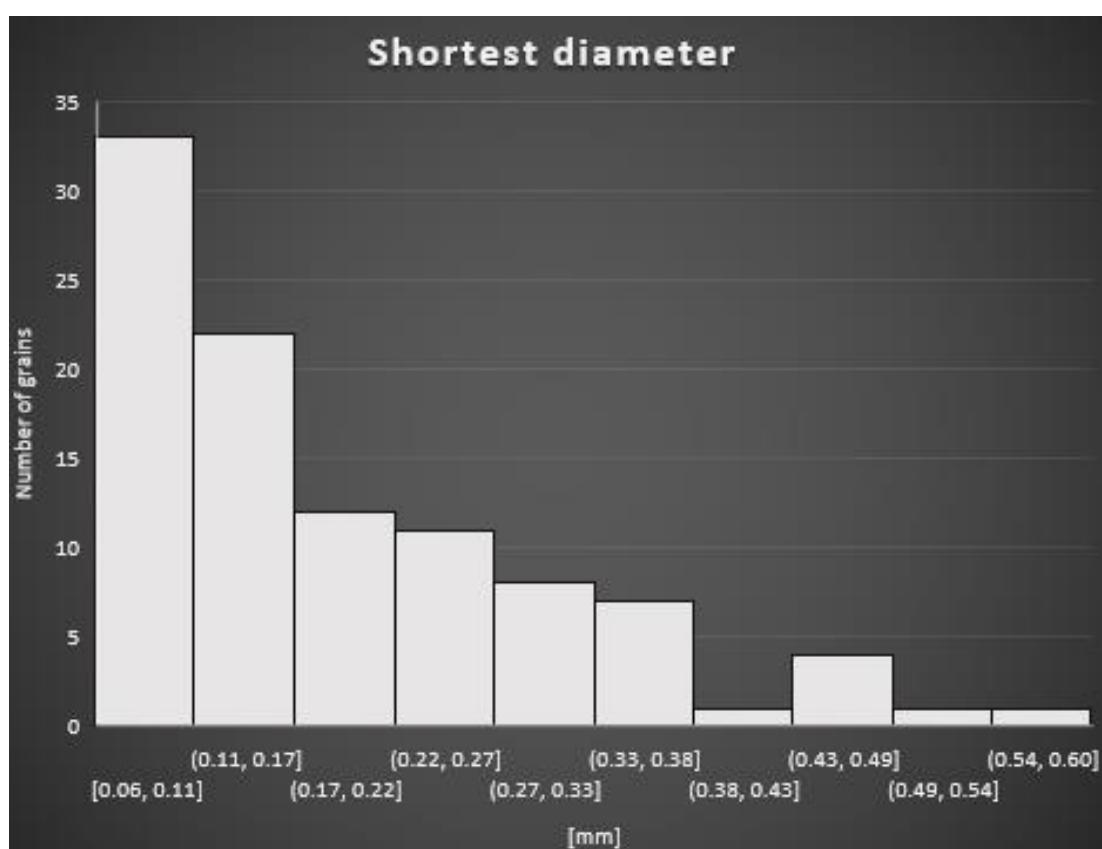
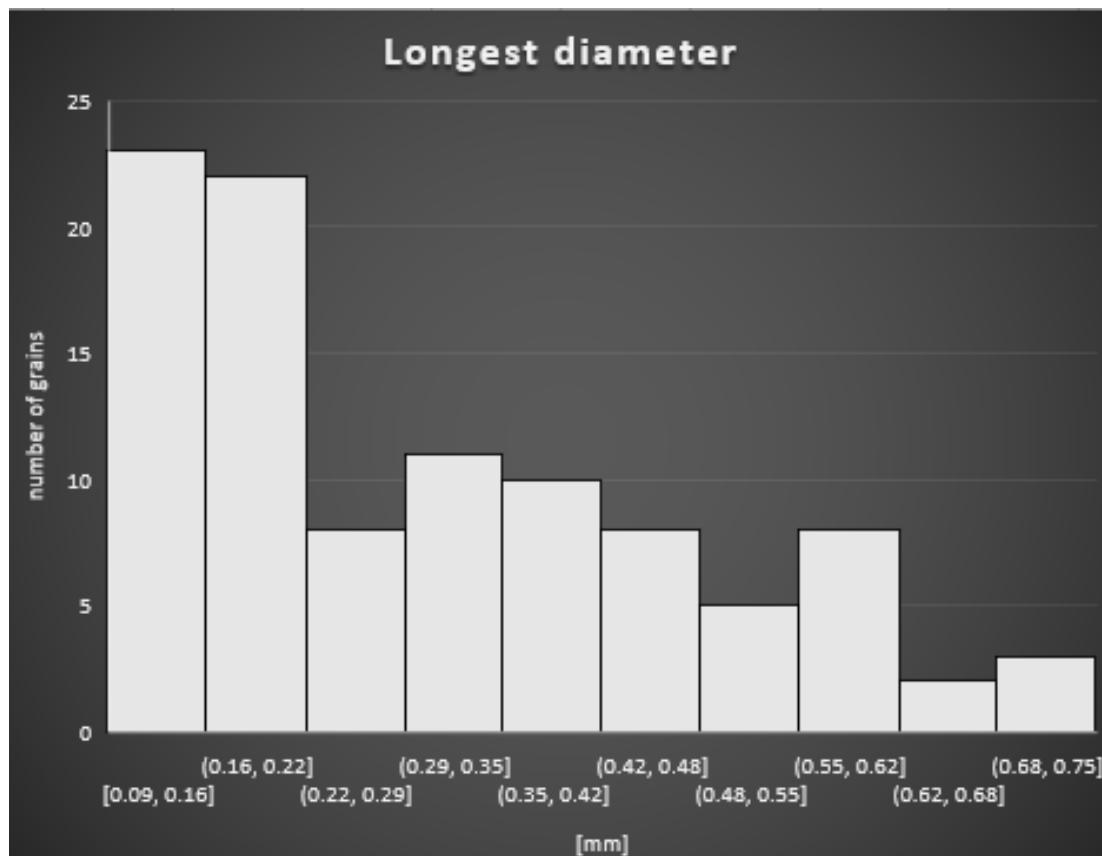


Figure A14 - Histograms for long (top) and short (bottom) diameters measured for **Sand ESF**. Each plot has 100 measurements distributed in 10 intervals along the x-axis and the number of sand-grains along the y-axis

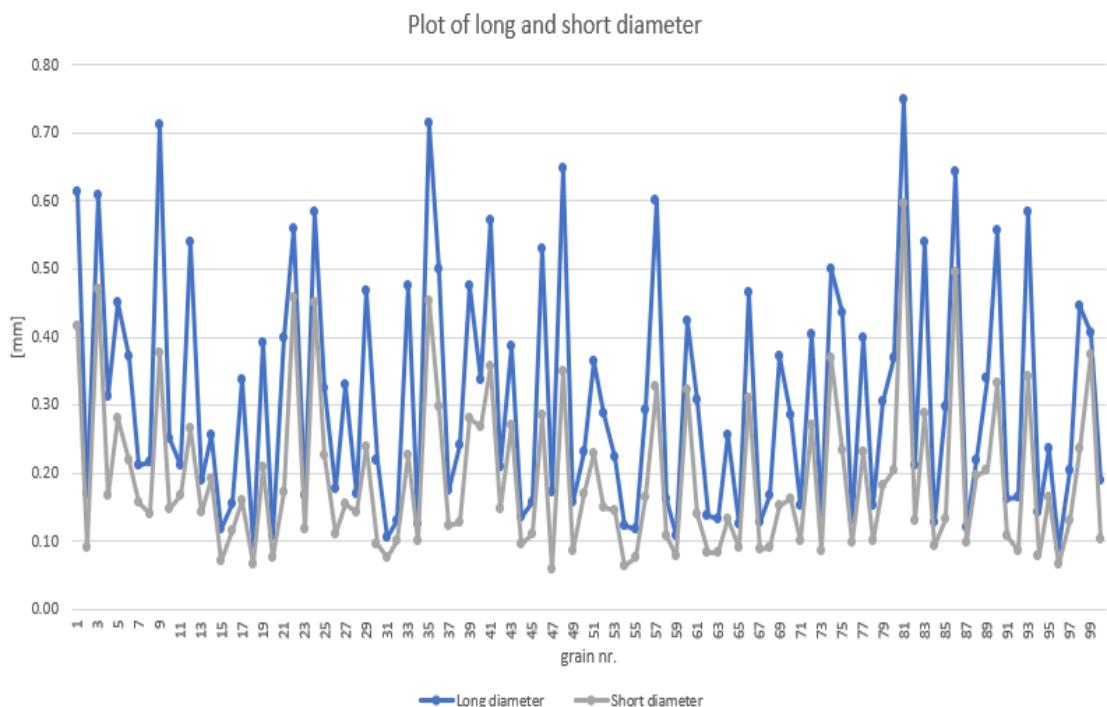


Figure A15 – Graphs for long (top, blue) and short (bottom, gray) grain diameters measured for **Sand ESF** using image in Figure A13. Each graph has 100 measurements.

Facies 6 – Sand G

Grain size distribution

- Sieve sizes used: 2.0 mm – 2.8 mm

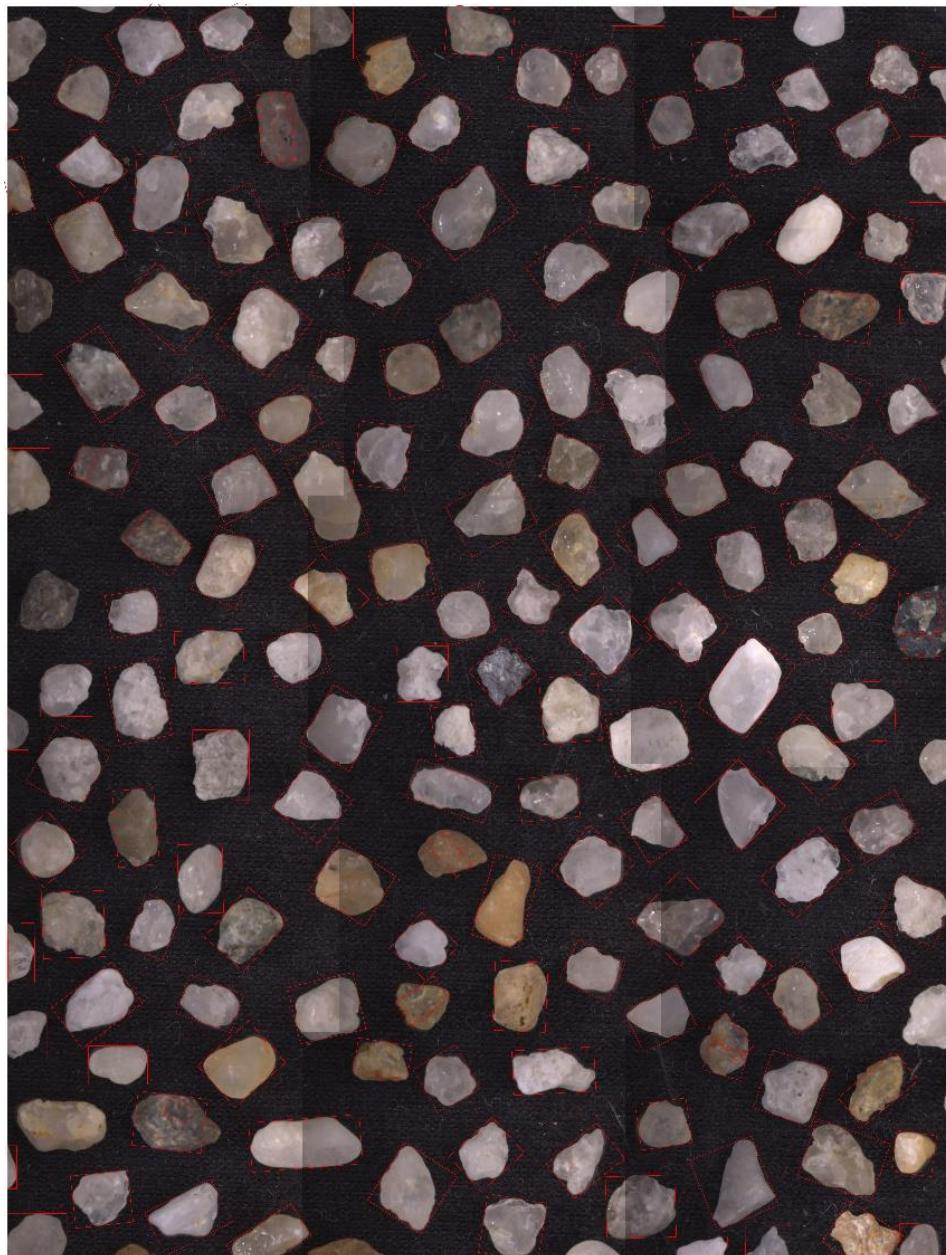


Figure A16 - A microscope image of **Sand G** and measured values for each of the 168 sand grains. The image shows the grains bounded by rectangles. To consider the rotation of the grains, the rectangles are drawn with minimum area. Grains that have similar color (dark grey and black) to the background on a binary image were not segmented. Grains with a partially similar color to the background was only partially segmented so that the calculated lengths for these grains are less than the actual length. Grains located close to the border, only partially visible in the image, are included in the analysis and contribute to the population of very small lengths. This will be corrected in future analysis.

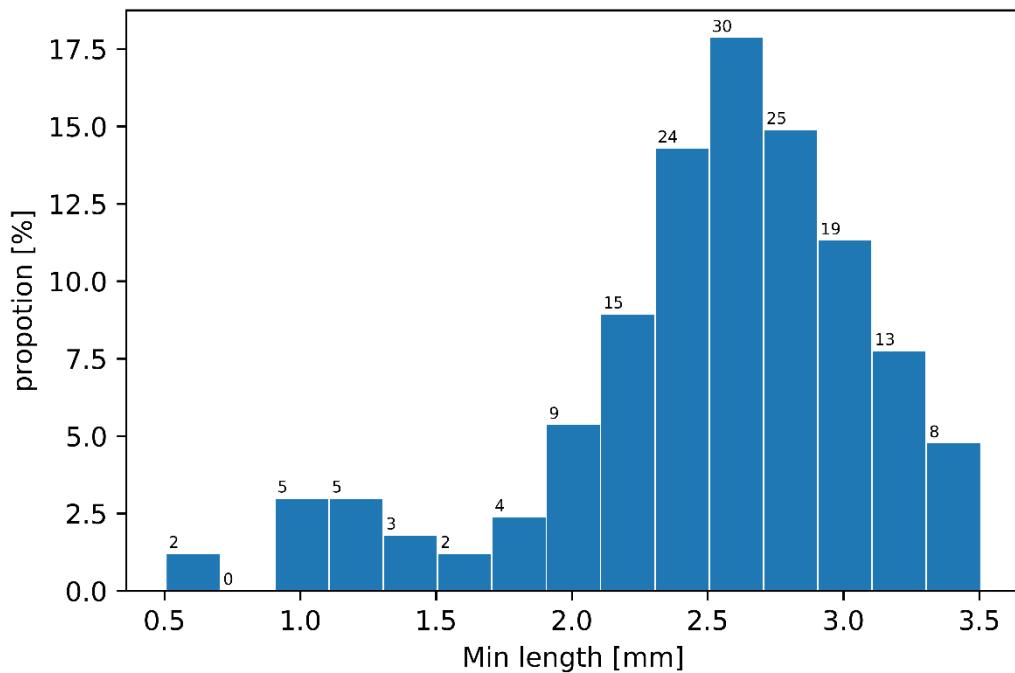
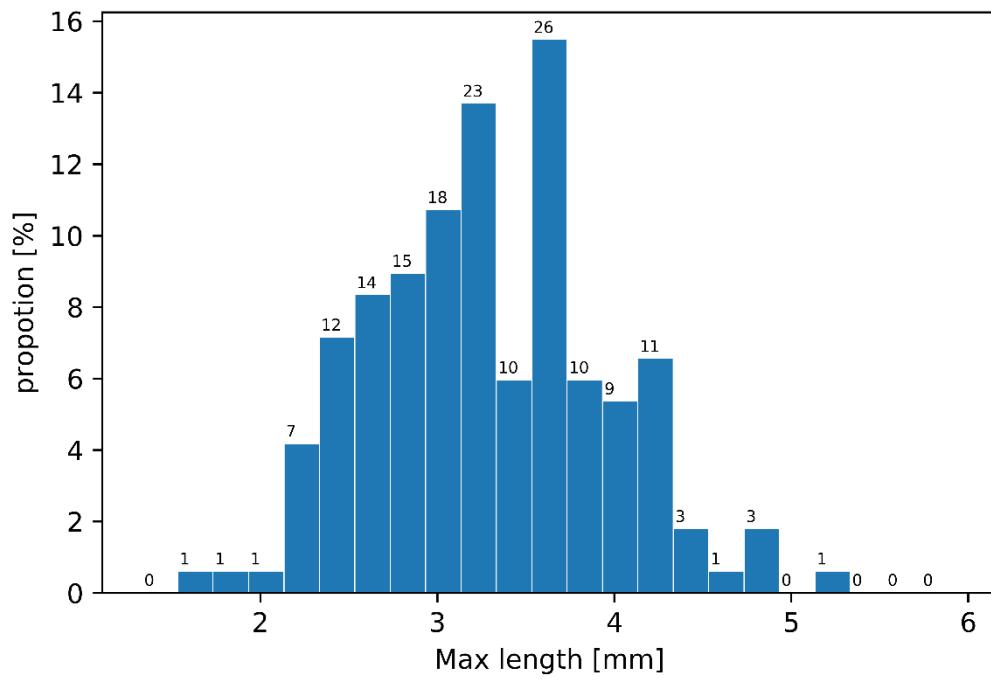


Figure A17 - Histograms for long (top) and short (bottom) grain lengths measured for **Sand G**. Number of grains in each population is indicated.

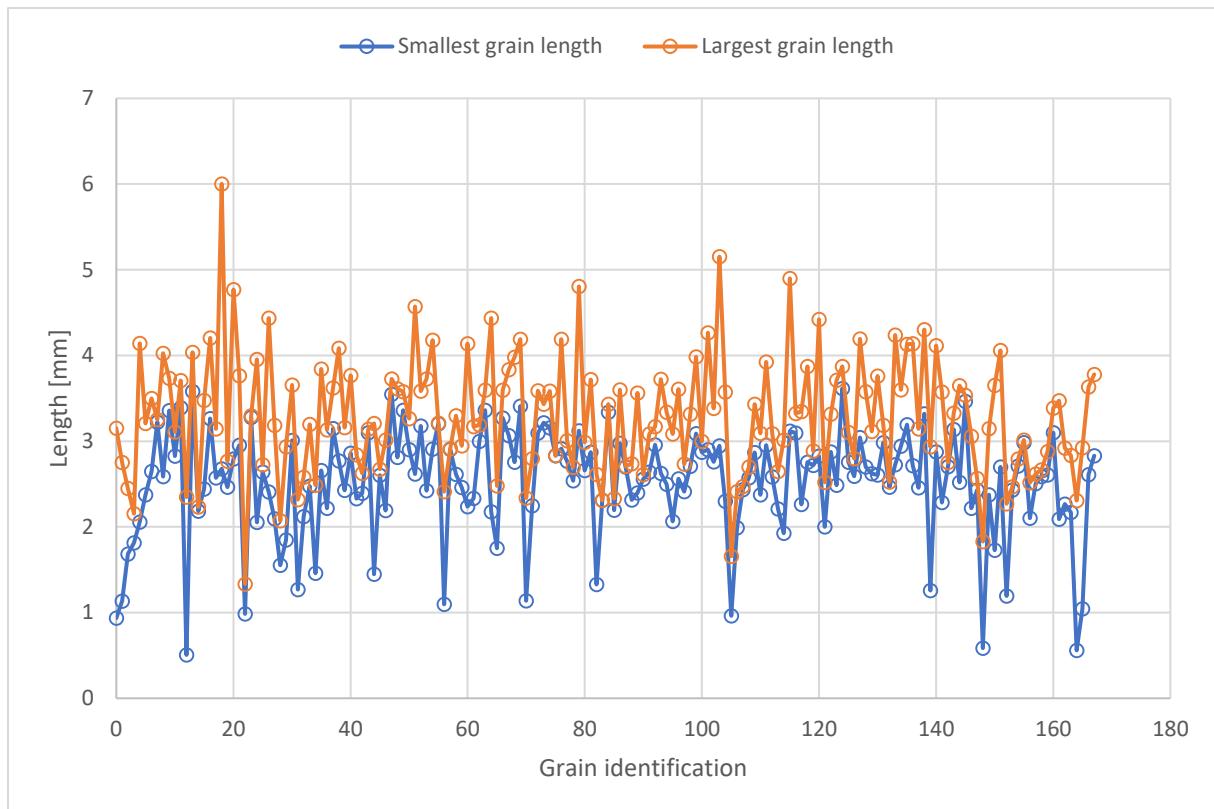


Figure A18 – Graphs for long (top, blue) and short (bottom, gray) grain diameters measured for **Sand G** using image in Figure A4. Each graph has 168 measurements.

Appendix B: Well test results

This appendix shows images from three well-test (1 in upper injection port, 2 in lower injection point), where aqueous phases with different colors (density of blue tracer was slightly higher than clear tracer – DI water) were injected. A time-lapse video and corresponding pressure responses are available at <https://github.com/fluidflower>.

The injection rate was 2250 ml/h, kept for 30 min, then 0 ml/h for 30 min. A total of 3 injection and pauses.

First injection: injection port (17,7)

Injection start: 5:19pm day 1

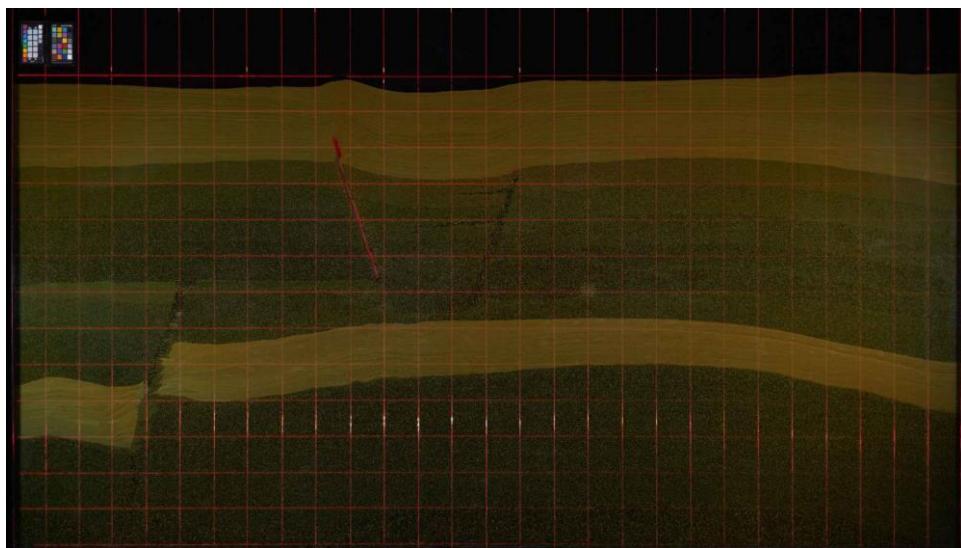


Figure B1: Start of injection

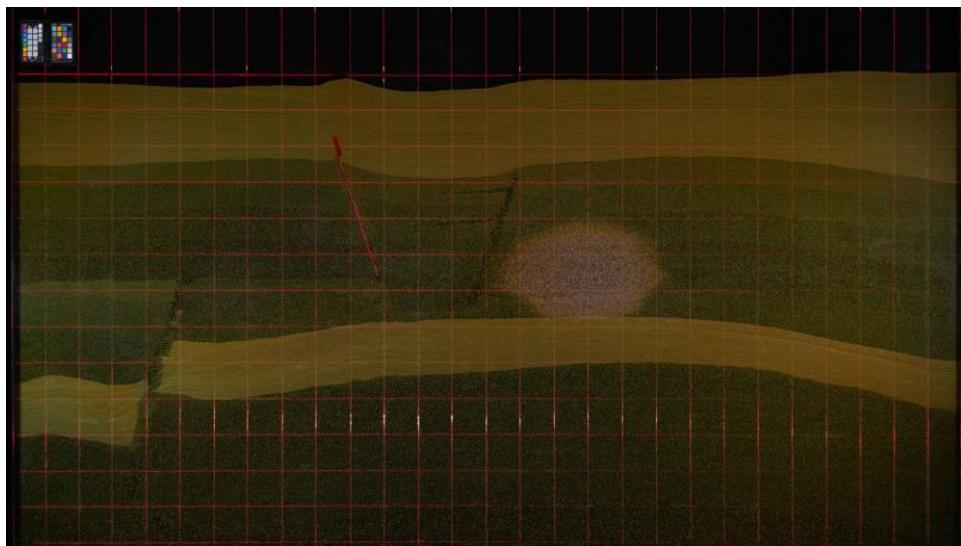


Figure B2: First pulse finished - end of first pulse: 5:49pm (30 min ± 3 sec) day 1

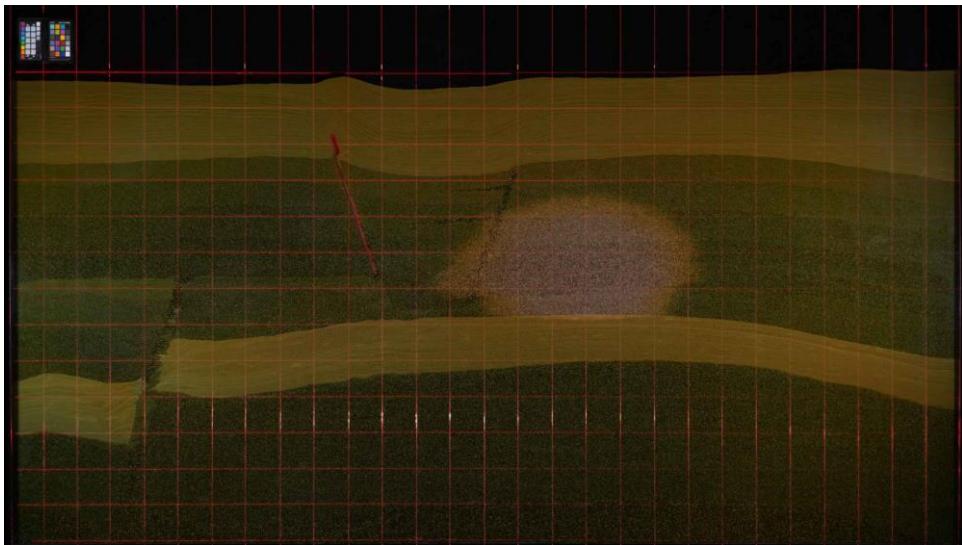


Figure B3: Second pulse finished – start of injection of 2nd pulse 18:20pm day 1

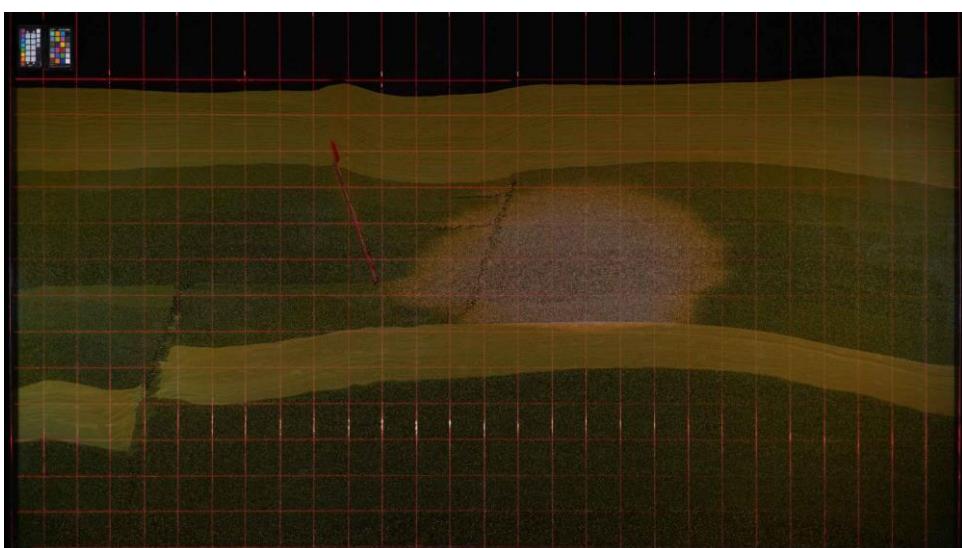


Figure B4: Third pulse finished – start

Second injection: injection port (9,3)

Same injection sequence as above.

Start injection 8:52 pm day 1, no delay between pulses.

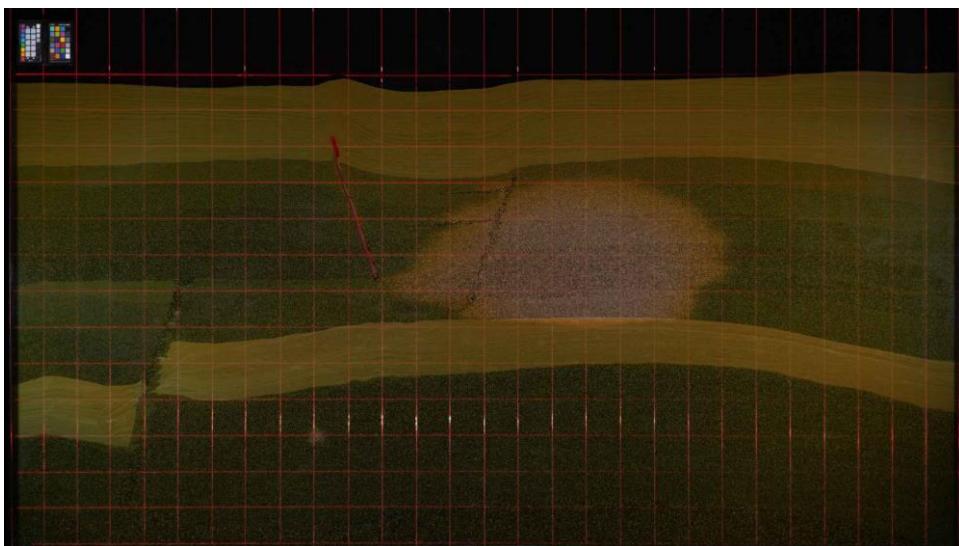


Figure B5: Start of injection

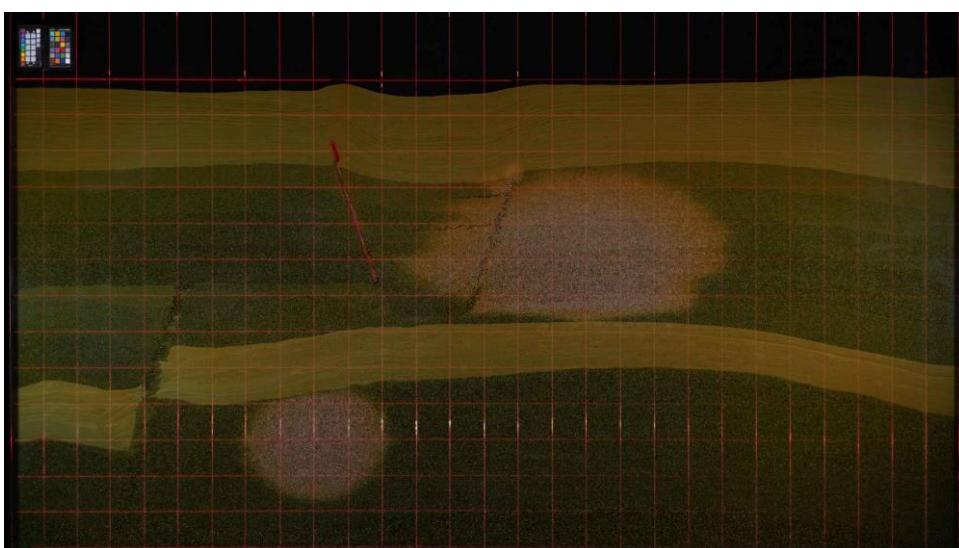


Figure B6: First pulse finished

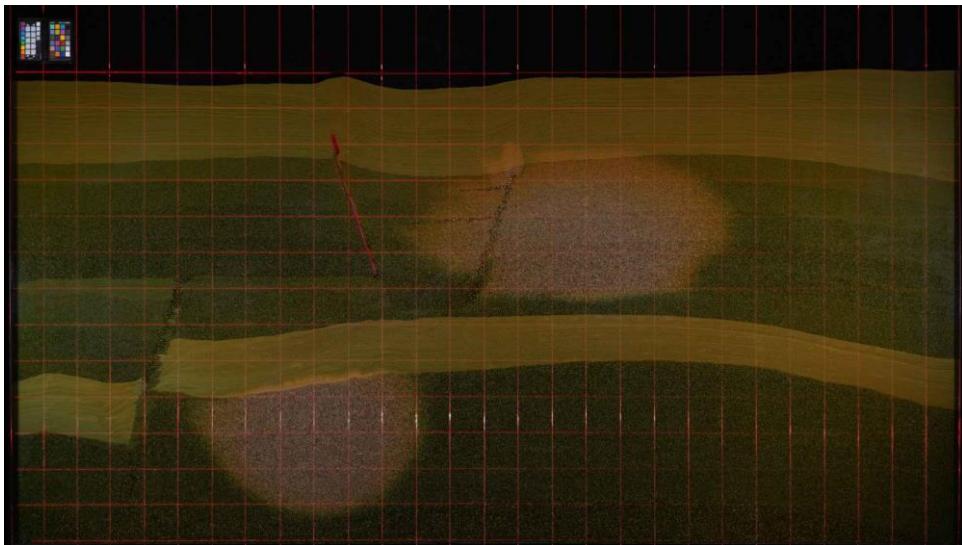


Figure B7: Second pulse finished

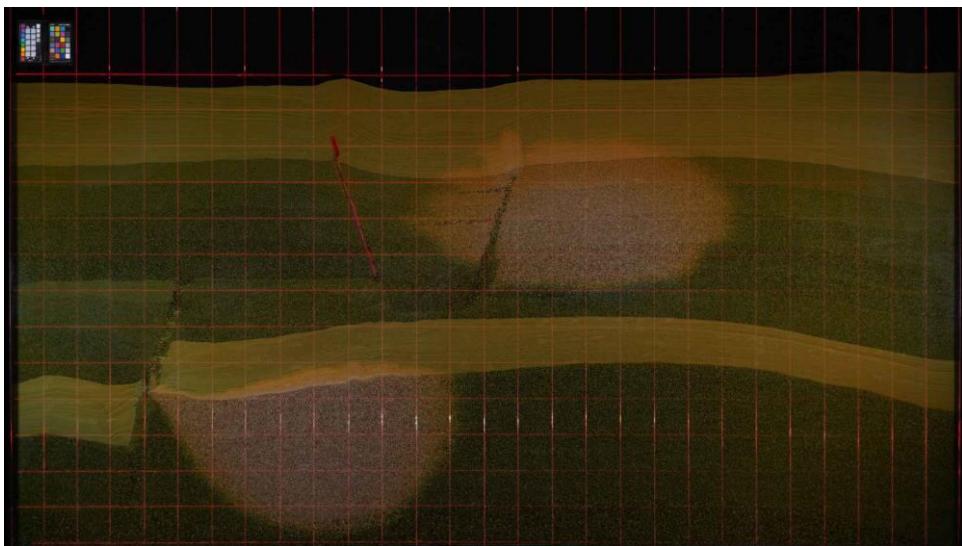


Figure B8: Third pulse finished

Third injection: 2nd injection port (9,3)

Same injection sequence as above.

Injection start 2:42pm day 2

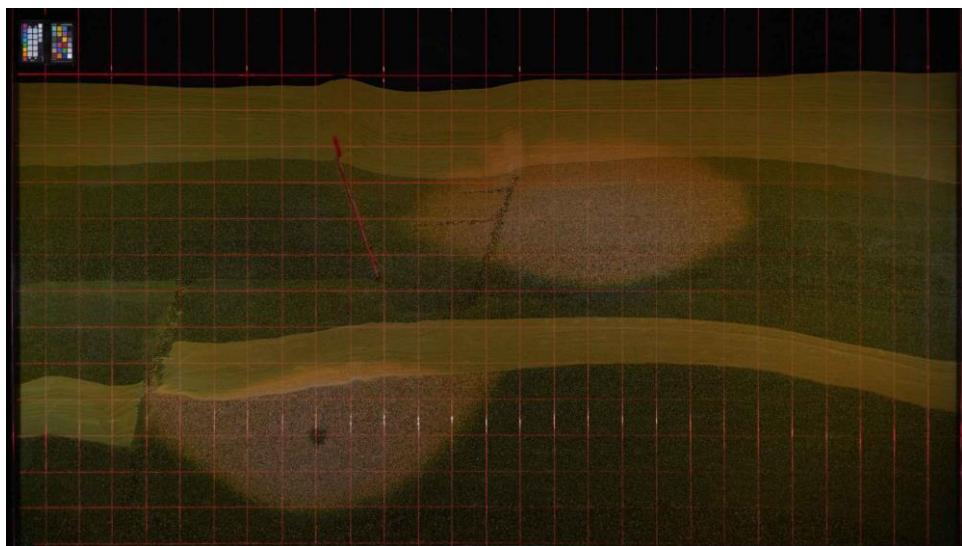


Figure B9: Start of injection

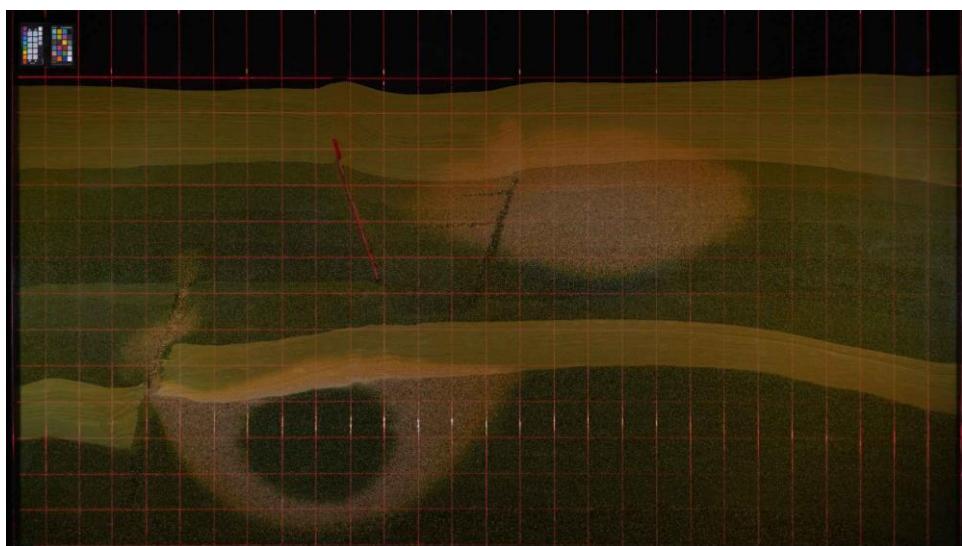


Figure B10: First pulse finished

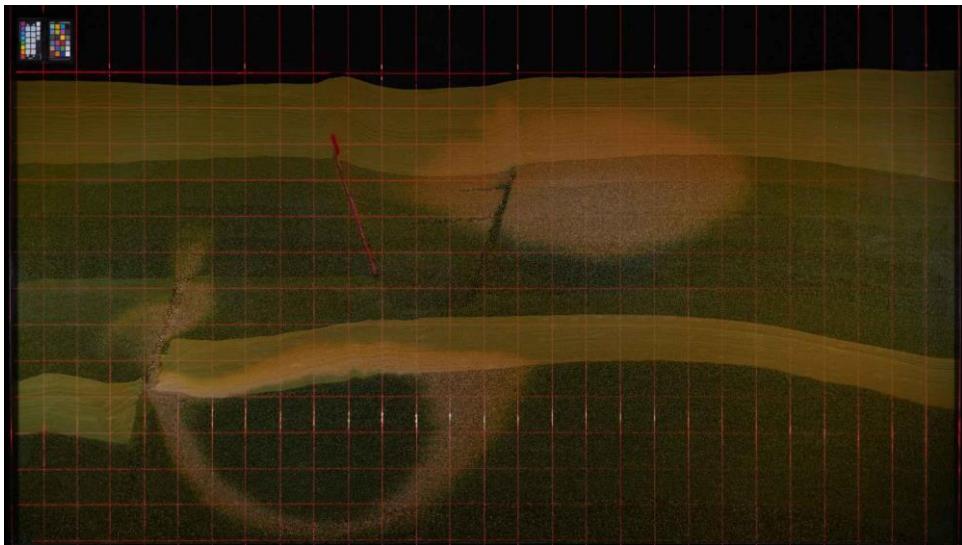


Figure B11: Second pulse finished

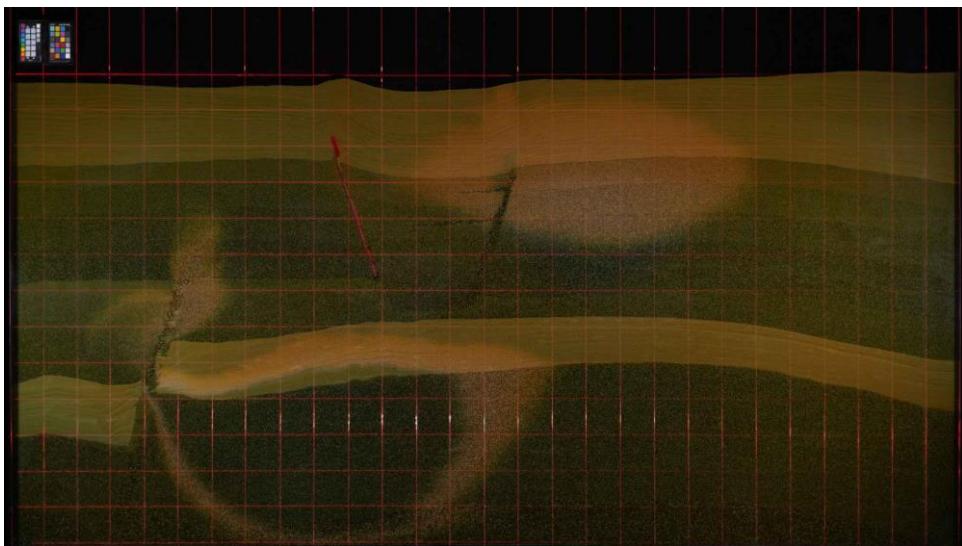


Figure B12: Third pulse finished